

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 13:47:50 ; Search time 2199.24 Seconds  
(without alignments)  
1283.613 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHASLGDSSETLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues 4109280  
Total number of hits satisfying chosen parameters:  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_173955\_334/app\_query.fasta\_1.462  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn\_1\_1758\_@runat\_05052003\_173955\_334 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	532	100.0	337	6	AX321911	AX321911 Sequence
2	527	99.1	5981	6	AX321910	AX321910 Sequence
3	513	96.4	2239	6	AX321909	AX321909 Sequence
C 4	322	60.5	161280	9	AC093903	AC093903 Homo sapi
5	221	41.5	575	6	AX341060	AX341060 Sequence
C 6	221	41.5	596	6	AX351341	AX351341 Sequence
7	221	41.5	1683	9	AB042201	AB042201 Homo sapi
8	221	41.5	1861	9	AB026891	AB026891 Homo sapi
9	221	41.5	1874	9	AF200708	AF200708 Homo sapi
10	221	41.5	2000	9	AB040875	AB040875 Homo sapi
11	221	41.5	2155	9	BC012087	BC012087 Homo sapi
12	221	41.5	2482	9	AF252872	AF252872 Homo sapi
13	210.5	39.6	3144	9	HSA277882	AJ277882 Homo sapi
14	144.5	27.2	2216	10	AB022345	AB022345 Mus muscu
15	144.5	27.2	2316	10	AB03765S01	AB037650 Mus muscu
16	131	24.6	96640	2	AC106184	AC106184 Rattus no
C 17	131	24.6	183949	2	AC118181	AC118181 Rattus no
C 18	131	24.6	202985	2	AC127139	AC127139 Rattus no
19	114	21.4	189747	2	AC087791	AC087791 Homo sapi
20	112.5	21.1	190815	9	AC007066	AC007066 Homo sapi
21	112.5	21.1	190824	2	AL731645	AL731645 Homo sapi
22	112	21.1	155025	9	AC034246	AC034246 Homo sapi
23	112	21.1	155365	9	AC107970	AC107970 Homo sapi
C 24	112	21.1	168720	2	AC078781	AC078781 Homo sapi
25	112	21.1	194334	2	AC046140	AC046140 Homo sapi
C 26	112	21.1	197110	9	AC104306	AC104306 Homo sapi
C 27	111.5	21.0	182509	9	AC112211	AC112211 Homo sapi
C 28	111	20.9	107795	9	AL662873	AL662873 Human DNA
29	111	20.9	149850	9	AP001622	AP001622 Homo sapi
30	111	20.9	340000	9	AP001746	AP001746 Homo sapi
31	110.5	20.8	110000	2	AC080008_0	AC080008 Homo sapi
C 32	110.5	20.8	215077	9	AC112504	AC112504 Homo sapi
C 33	110	20.7	118819	9	AC011491	AC011491 Homo sapi
34	109.5	20.6	39958	9	AC011542	AC011542 Homo sapi
35	109.5	20.6	41716	9	AC011522	AC011522 Homo sapi
C 36	109	20.5	41572	9	AC006132	AC006132 Homo sapi
C 37	109	20.5	77322	2	AC021334	AC021334 Homo sapi
C 38	109	20.5	127361	9	AC007283	AC007283 Homo sapi
C 39	109	20.5	145414	9	HSA392M18	AL121897 Human DNA
C 40	109	20.5	179216	2	AC074016	AC074016 Homo sapi
C 41	109	20.5	184063	9	AC023481	AC023481 Homo sapi
42	108.5	20.4	124086	9	AC008882	AC008882 Homo sapi
C 43	108.5	20.4	127171	2	AC008777	AC008777 Homo sapi
C 44	108.5	20.4	145264	9	AC016601	AC016601 Homo sapi
C 45	108.5	20.4	150296	9	HUU91326	U91326 Homo sapien

ALIGNMENTS

```
AX321911
LOCUS AX321911 337 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 442 from Patent WO0172295.
ACCESSION AX321911
VERSION AX321911.1 GI:17906521
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,
Indirias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,
Mannion,J. and Kalos,M.D.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0172295-A 442 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
BASE COUNT 103 a 60 c 93 g 81 t
ORIGIN
Alignment Scores: 1.72e-55 Length: 337
Pred. No.: 532.00 Matches: 97
Score: 100.00%
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Mismatches: 0
Indels: 0
DB: 6
US-09-854-133-586 (1-97) x AX321911 (1-337)
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20
Db 5 GAGGTTGAAGTGAGCAGAGATCATGCCAGCTGGGTGACAGTGTCTCTCAACA 64
QY 21 GluLeuArgLysLysGluArgLysLysArgGluArgLysPheGlnAlaAsnCysGly 40
Db 65 GAATTAAGGAAAAAGAAAGAAAGAGAGAGAGAGAAATCCAGGCCAATTGTGGC 124
QY 41 IleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGln 60
Db 125 ATAGATTTTATCATATCTGGATTTTGGATTTCTTTGTTCTCATCTGATTTCAG 184
QY 61 GluSerLeuLeuCysProSerProSerProLysGluValThrCysArgGluMetLeuThrGly 80
Db 185 GAAAGCCTGTTGTCTCCACCATCTCCAAAGAGAGGTACCTGCAGGAAATGTTAACGGGA 244
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97
Db 245 GGCTGCCTTCCCTGGGCAACAAGGAGCCACCTGGCAGGAGAAAGTGCAGC 295
RESULT 2
AX321910
LOCUS AX321910 5981 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 441 from Patent WO0172295.
ACCESSION AX321910
VERSION AX321910.1 GI:17906518
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,
Indirias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,
Mannion,J. and Kalos,M.D.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0172295-A 441 04-OCT-2001;
```

```
AX321911
LOCUS AX321911 5981 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 440 from Patent WO0172295.
ACCESSION AX321909
VERSION AX321909.1 GI:17906515
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,
Indirias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,
Mannion,J. and Kalos,M.D.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0172295-A 440 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
BASE COUNT 619 a 444 c 493 g 683 t
ORIGIN
Alignment Scores: 2.97e-52 Length: 2239
Pred. No.: 513.00 Matches: 97
Score: 84.35%
Percent Similarity: 84.35%
Best Local Similarity: 84.35%
Query Match: 96.43%
Indels: 6
DB: 6
US-09-854-133-586 (1-97) x AX321909 (1-2239)
QY 2 ValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThrGlu 21
Db 3 GTTGAAGTGAGCAGAGATCATGCCAGCTGGGTGACAGTGTCTCTCAACAGAA 62
QY 22 LeuArgLysLysGluArgLysLysArgGluArgLysPheGlnAlaAsnCysGly 41
Db 63 TTAAGGAAAAAGAAAGAAAGAGAGAGAGAAATCCAGGCCAATTGTGGCATA 122
QY 42 AspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGlnGlu 61
Db 123 GATTTATCATATCTGGATTTTGGATTTCTTTGTTCTCATCTGATTCAGGAA 182
QY 62 SerLeuLeuCysProSerProSerProLysGluValThrCysArgGluMetLeuThrGly 81
Db 183 AGCCTGTTGTCTCCACCATCTCCAAAGAGAGGTACCTGCAGGAAATGTTAACGGGAGGC 242
QY 82 CysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97
Db 243 TGCCTTCCCTGGGCAACACAGGAGCCACCTGGGAGGAGAAAGTGCAGC 290
RESULT 3
AX321909
LOCUS AX321909 2239 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 440 from Patent WO0172295.
ACCESSION AX321909
VERSION AX321909.1 GI:17906515
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,
Indirias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,
Mannion,J. and Kalos,M.D.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0172295-A 440 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
BASE COUNT 619 a 444 c 493 g 683 t
ORIGIN
Alignment Scores: 2.97e-52 Length: 2239
Pred. No.: 513.00 Matches: 97
Score: 84.35%
Percent Similarity: 84.35%
Best Local Similarity: 84.35%
Query Match: 96.43%
Indels: 6
DB: 6
US-09-854-133-586 (1-97) x AX321909 (1-2239)
```



QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
Db 2 GAGTTGAAGTGAGCAGAGATCATGCCAGCCTGGTGACAGTGAGACTCTGTCTCAACA 61  
QY 21 GluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 40  
Db 62 GAATTAAGGAAAAAAGAAAGAAAGAGAGAGAGAGAGAAATCCAGGCCAATTTGTGGC 121  
QY 41 ileAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGln 60  
Db 122 ATAGATTTTATCATATCTGGATTTTGGATTTCTTTTCTCATCATCTGGATTCAG 181  
QY 61 GluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
Db 182 GAAAGCCTGTTGTCTCCACCATCTCCAAAGGAGGTTACCTGCAGGGAAATGTTAACGGGA 241  
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArg----- 93  
Db 242 GGCTGCCTTCCCTGGCAACAAGAGAGCCACCTGGCGAGGCGCTTTTCAGGAAGAGACG 301  
QY 94 -----ArgLysCysSer 97  
Db 302 CCTTTTCAGGAAGAGACGCGCTTTTCAGGAAGAGAGAGAAAGTGCAGC 346  
RESULT 4  
AC093903/c  
LOCUS AC093903 161280 bp DNA linear PRI 01-MAR-2002  
DEFINITION Homo sapiens BAC clone RP11-733C7 from 4, complete sequence.  
ACCESSION AC093903 AC055827  
VERSION AC093903.3 GI:15920156  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 161280)  
Sulston, J.E. and Waterston, R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
99063792  
PUBMED 9847074  
2 (bases 1 to 161280)  
Radionenko, M. and Kozlowicz, A.  
The sequence of Homo sapiens BAC clone RP11-733C7  
Unpublished (2001)  
3 (bases 1 to 161280)  
Waterston, R.H.  
Direct Submission  
Submitted (10-SEP-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 161280)  
Waterston, R.H.  
Direct Submission  
Submitted (04-OCT-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
5 (bases 1 to 161280)  
Waterston, R.  
Direct Submission  
Submitted (01-MAR-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Oct 4, 2001 this sequence version replaced gi:15625016.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0733C07  
Drafting Center: WIBR  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the right is RP11-310A13. Actual start of this clone is at base position 1 of RP11-733C7; actual end is at base position 161280 of RP11-733C7.

Data from AC009792 was used to finish the clone, AC055827.

The sequence of AC055827 has been incorporated into AC093903.

FEATURES  
source  
1. .161280  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-733C7"  
/clone\_lib="RPCI-11"  
148. .304  
/rpt\_family="L1"  
repeat\_region  
306. .406  
/rpt\_family=" (GGAA)n"  
repeat\_region  
471. .525  
/rpt\_family="AT-rich"  
repeat\_region  
530. .782  
/rpt\_family="L1"  
repeat\_region  
685. .712  
/rpt\_family=" (TA)n"  
repeat\_region  
1383. .1472  
/rpt\_family="GA-rich"  
repeat\_region  
1697. .1836  
/rpt\_family="L1"  
repeat\_region  
1837. .1956  
/rpt\_family="L1"  
repeat\_region  
3005. .3221  
/rpt\_family="MIR"  
repeat\_region  
4013. .4321  
/rpt\_family="Alu"  
repeat\_region  
4295. .4325  
/rpt\_family=" (A)n"  
repeat\_region  
5625. .5960  
/rpt\_family="L1"  
repeat\_region  
5938. .5982  
/rpt\_family="A-rich"  
repeat\_region  
6062. .6195

```
repeat_region /rpt_family="MIR"
6388. .6462
repeat_region /rpt_family="L2"
6546. .6868
misc_feature /rpt_family="L1"
6866. .7318
repeat_region /note="similar to EST C06175 (NID:g1502951)"
6866. .6898
misc_feature /rpt_family="AT_rich"
7134. .7385
misc_feature /note="similar to EST BG989897 (NID:g14393967)"
7777. .7918
repeat_region /note="similar to EST BG388107 (NID:g13281553)"
8483. .8522
repeat_region /rpt_family="AT_rich"
8721. .9018
repeat_region /rpt_family="Alu"
9459. .9741
repeat_region /rpt_family="L2"
10217. .10246
repeat_region /rpt_family="AT_rich"
10304. .10385
repeat_region /rpt_family="MIR"
10835. .10878
misc_feature /rpt_family="AT_rich"
11861. .11987
misc_feature /note="similar to EST BG388107 (NID:g13281553)"
11873. .11987
repeat_region /note="similar to EST BF286052 (NID:g11217122)"
12516. .12539
repeat_region /rpt_family="(T)n"
12929. .13232
repeat_region /rpt_family="Alu"
13237. .13274
repeat_region /rpt_family="(TA)n"
13275. .13304
repeat_region /rpt_family="(T)n"
13696. .13838
repeat_region /rpt_family="MIR"
14127. .14153
repeat_region /rpt_family="AT_rich"
14132. .14436
repeat_region /rpt_family="Alu"
15823. .16117
misc_feature /rpt_family="Alu"
17329. .17894
misc_feature /note="similar to EST BG388107 (NID:g13281553)"
17329. .17618
misc_feature /note="similar to EST BF286052 (NID:g11217122)"
17489. .17611
misc_feature /note="similar to EST AT133891 (NID:g4029010)"
17548. .17869
misc_feature /note="similar to EST BF095483 (NID:g10901193)"
17772. .18329
repeat_region /note="similar to EST BG181259 (NID:g13703031)"
17845. .17869
repeat_region /rpt_family="(CTG)n"
18792. .18828
repeat_region /rpt_family="AT_rich"
20734. .20784
repeat_region /rpt_family="MER1_type"
20832. .20944
repeat_region /rpt_family="MER1_type"
20948. .21964
repeat_region /rpt_family="L1"
21964. .21986
repeat_region /rpt_family="AT_rich"
21965. .22276
repeat_region /rpt_family="Alu"
22277. .22323
repeat_region /rpt_family="L1"
22341. .22681
repeat_region /rpt_family="L1"
```

```
repeat_region 22685. .22824
/rpt_family="Alu"
repeat_region 22805. .22824
/rpt_family="(A)n"
repeat_region 22826. .23575
/rpt_family="L1"
repeat_region 23591. .23735
/rpt_family="MER1_type"

Alignment Scores:
Pred. No.: 5.44e-27 Length: 161280
Score: 322.00 Matches: 60
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 60.53% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-586 (1-97) x AC093903 (1-161280)
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20
|||||
Db 27755 GAGTTGAAGTGAGCAGATCATGCCAGCTGGTGACAGTGAGACTCTGTCTCAACA 27696
QY 21 GluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 40
|||||
Db 27695 GAATTAAGGAAAAAGAAAGAAAGAAAGAGAGAGAGAGAGAAATCCAGGCCAATTGTGGC 27636
QY 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGln 60
|||||
Db 27635 ATAGATTATCATATCTCGGATTTTGGATTCTTTGTTTCTCATCACTGGATTGAG 27576

RESULT 5
AX341060
LOCUS AX341060
DEFINITION Sequence 1307 from Patent WO0196388.
ACCESSION AX341060
VERSION AX341060.1 GI:18137042
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Jiang, Y., Harlocker, S.L. and Secrist, H.
TITLE Compositions and methods for the therapy and diagnosis of colon
cancer
JOURNAL Patent: WO 0196388-A 1307 20-DEC-2001;
CORIXA CORPORATION (US)
FEATURES
Source
1. .575
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 151 a 122 c 153 g 141 t 8 others
ORIGIN

Alignment Scores:
Pred. No.: 1.6e-17 Length: 575
Score: 221.00 Matches: 47
Percent Similarity: 70.00% Conservative: 9
Best Local Similarity: 58.75% Mismatches: 18
Query Match: 41.54% Indels: 6
DB: 6 Gaps: 2

US-09-854-133-586 (1-97) x AX341060 (1-575)
QY 19 GlnThrGluLeuArgLysLysGluArgLys---LysLysArgGluArgLysPheGlnAla 37
|||||
Db 142 CAGTCTGAAAGCAGAGAGAGACATCATCATAGTACAGTACACCAAGAGACACCAAGTGAAGT 201
QY 38 AsnCysGlyIleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHis 57
|||||
Db 202 TTTGTTTCTTCTCCCTCTGTTTATTTTCCCTCCCGTGTGTCCTACTATGTCATCA----- 255
```

QY 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77  
Db 256 -----GAAAGCCTGTTGTGTCACCATCTCCAAAGGAGGTTACTGTCAGGAAATG 306  
QY 78 LeuThrGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97  
Db 307 TTAACGGGAGGCTGCCTTCCCTGGGCAACAAGGAGCCACCTGGGAGGAGAAAGTGCAGC 366  
RESULT 6  
AX351341/c  
LOCUS AX351341 596 bp DNA linear PAT 06-FEB-2002  
DEFINITION Sequence 88 from Patent WO0196390.  
ACCESSION AX351341  
VERSION AX351341.1 GI:18616688  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Jiang,Y., Hepler,W.T., Clapper,J.D., Wang,A. and Secrist,H.  
TITLE Compositions and methods for the therapy and diagnosis of colon cancer  
JOURNAL Patent: WO 0196390-A 88 20-DEC-2001;  
CORIXA CORPORATION (US)  
FEATURES Location/Qualifiers  
source 1..596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 160 a 151 c 126 g 158 t 1 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.67e-17 Length: 596  
Score: 221.00 Matches: 47  
Percent Similarity: 70.00% Conservative: 9  
Best Local Similarity: 58.75% Mismatches: 18  
Query Match: 41.54% Indels: 6  
DB: 6 Gaps: 2  
US-09-854-133-586 (1-97) x AX351341 (1-596)  
QY 19 GlnThrGluLeuArgLysLysGluArgLys---LysLysArgGluArgLysPheGlnAla 37  
Db 481 CAGTCTGAAAGCAGAGGAGACATCGATCAGTAACACCAAGAGACACCAAGTTGAAAGT 422  
QY 38 AsnCysGlyIleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisHis 57  
Db 421 TTTGTTTCTTCTCTCTGTTTATTTTCCCGGTGTCTCCCTACTATGGTCA----- 368  
QY 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77  
Db 367 -----GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACCTGCAGGAAATG 317  
QY 78 LeuThrGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97  
Db 316 TTAACGGGAGGCTGCCTTCCCTGGGCAACAAGGAGCCACCTGGGAGGAGAAAGTGCAGC 257  
RESULT 7  
AB042201  
LOCUS AB042201 1683 bp DNA linear PRI 10-FEB-2001  
DEFINITION Homo sapiens xCT gene for cystine/glutamate transporter, exon 1,  
partial cds.  
ACCESSION AB042201  
VERSION AB042201.1 GI:7670267  
KEYWORDS cystine/glutamate transporter.  
SOURCE Homo sapiens DNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Sato,H., Tamba,M., Kuriyama-Matsumura,K., Okuno,S. and Bannai,S.  
TITLE Molecular cloning and expression of human xCT, the light chain of

amino acid transport system xc-  
Antioxidants and Redox Signaling 2, 665-671 (2000)  
2 (bases 1 to 1683)  
Sato,H. and Bannai,S.  
Direct Submission  
Submitted (24-APR-2000) Hideyo Sato, University of Tsukuba,  
Institute of Basic Medical Sciences; Tennodai 1-1-1, Tsukuba,  
Ibaraki 305-8575, Japan (E-mail:hideyo-s@md.tsukuba.ac.jp,  
Tel:81-298-53-3282, Fax:81-298-53-3039)  
Location/Qualifiers  
1..1683  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1001..1683  
/gene="xCT"  
1001..1555  
/gene="xCT"  
/product="cystine/glutamate transporter"  
/number=1  
1279..>1555  
/gene="xCT"  
/codon\_start=1  
/product="cystine/glutamate transporter"  
/protein\_id="BAA94999.1"  
/db\_xref="GI:7670268"  
/translation="MVRKPVVSTISKGYLQGVNVRPLPSLGNKEPPGQEKVQLKRV  
TLRGVSIIGTIIGAGIFISPKVLQNTGSGVMSLTITWTCVLSLF"  
1556..>1683  
/gene="xCT"  
/number=1  
BASE COUNT 495 a 287 c 370 g 531 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 5.41e-17 Length: 1683  
Score: 221.00 Matches: 47  
Percent Similarity: 70.00% Conservative: 9  
Best Local Similarity: 58.75% Mismatches: 18  
Query Match: 41.54% Indels: 6  
DB: 9 Gaps: 2  
US-09-854-133-586 (1-97) x AB042201 (1-1683)  
QY 19 GlnThrGluLeuArgLysLysGluArgLys---LysLysArgGluArgLysPheGlnAla 37  
Db 1172 CAGTCTGAAAGCAGAGGAGACATCGATCAGTAACACCAAGAGACACCAAGTTGAAAGT 1231  
QY 38 AsnCysGlyIleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisHis 57  
Db 1232 TTTGTTTCTTCTCTCTGTTTATTTTCCCGGTGTCTCCCTACTATGGTCA----- 1285  
QY 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77  
Db 1286 -----GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACCTGCAGGAAATG 1336  
QY 78 LeuThrGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97  
Db 1337 TTAACGGGAGGCTGCCTTCCCTGGGCAACAAGGAGCCACCTGGGAGGAGAAAGTGCAGC 1396  
RESULT 8  
AB026891  
LOCUS AB026891 1861 bp mRNA linear PRI 10-FEB-2001  
DEFINITION Homo sapiens mRNA for cystine/glutamate transporter, complete cds.  
ACCESSION AB026891  
VERSION AB026891.1 GI:5668544  
KEYWORDS cystine/glutamate transporter.  
SOURCE Homo sapiens fibroblast cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Sato,H., Tamba,M., Kuriyama-Matsumura,K., Okuno,S. and Bannai,S.  
TITLE Molecular cloning and expression of human xCT, the light chain of

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
amino acid transport system xc-  
Antioxidants and Redox Signaling 2, 665-671 (2000)  
2 (bases 1 to 1861)  
Sato,H. and Bannai,S.  
Direct Submission  
Submitted (30-APR-1999). Hideyo Sato, University of Tsukuba,  
Institute of Basic Medical Sciences; Tennodai 1-1-1, Tsukuba,  
Ibaraki 305-8575, Japan (E-mail:hideyo-s@md.tsukuba.ac.jp,  
Tel:81-298-53-3282, Fax:81-298-53-3039)

Location/Qualifiers	source
1. .1861	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/cell_type="fibroblast"	
1. .1861	
/gene="xCT"	
236. .1741	
/gene="xCT"	
/codon_start=1	
/product="cystine/glutamate transporter"	
/protein_id="BAA82628.1"	
/db_xref="GI:5668545"	
/translation="MVRKPVVSTISKGYLQGNVNGRLPSLGNKEPPGQEKVQLKRKV TLRGVSIIGTIIGAGFIIPKGVLTQNTSGVMSLTITWTCVGLSLFGALSYAELGT TIKKSGHYTYILEVFGPLPAFVRVWVELLIIRPAATAVISLAFGRYILEPFIQCEI PELAIKLITAVGITVVMVLMNSMSVSWSARIQIFLTFCKLTAILIIIVPGVMQLIKQT QNFKDAFSGRDSIIIRLPLAFYGYMYAYAGWFYLNFTVEEVENPEKTIPLAICISMAI VTIGVLTNVAYFTTINAEELLSNAVAVTFSERLIGNFSLAVPIFVALSCFGSMNGG VFVASRLFYVASREGHLP EILSMIHVRKHTPLPAVIVLHPLTMIMLFSGDLDSLNLFL SFARWLFIGLAVAGLIYLRKPCDMMHRPFKVPPLFIPALFSTCLEFMVALSLYSDPFT GIGFVITLTGVPAYLYFIWDKKPRWFRIMSEKITRTTLQILEVVPPEDKL"	
485 a 402 c 421 g 553 t	
BASE COUNT	
ORIGIN	

Alignment Scores:		
Pred. No.:	6.07e-17	1861
Score:	221.00	47
Percent Similarity:	70.00%	9
Best Local Similarity:	58.75%	18
Query Match:	41.54%	6
DB:	9	2
	Length:	
	Matches:	
	Conservative:	
	Mismatches:	
	Indels:	
	Gaps:	

US-09-854-133-586 (1-97) x AB026891 (1-1861)

```

QY      19 GlnThrGluLeuArgLysLysLysGluArgLys---LysLysArgGluArgLysPheGlnAla 37
        |||::|||   ||| ::: |||          |||:::   |||       |
DbB     129 CAGTCTGAAGCAGAGGAACATCGATCACGTAAACCACAAGACACCAAGTTGAAGT 188
                                     ::::||||:::
QY      38 AsnCysGlyIleAspPheIleilePheTrpIlePheTrpIleLeuLeuPheSerHisHis 57
                                   ::::||||:::
                                           |||:::||||
DbB    189 TTTGTTTTCTTTCCCTCTGTGTTTTATTTCCTCCCCGGTGTCCTACTATGGTCA----- 242

```

DY	58	T	r	p	i	e	g	l	n	g	l	s	e	r	l	e	u	c	y	s	p	r	o	s	e	r	p	r	o	l	y	s	g	l	u	v	a	l	t	h	r	c	y	s	a	r	g	l	u	m	e	t	77						
Db	243	-	-	-	-	-	-	-	-	-	G	A	A	G	C	T	G	T	G	T	G	T	C	C	A	C	C	A	T	C	T	C	C	A	A	G	G	A	G	G	T	T	A	C	T	G	C	A	G	G	A	A	T	G	293				
DY	78	L	e	u	t	h	r	g	l	y	G	l	y	C	ys	L	e	u	P	r	o	T	r	p	A	l	a	T	h	r	A	r	g	S	e	r	H	i	s	L	e	u	G	l	y	A	r	g	A	r	g	L	ys	C	ys	S	e	r	97
db	294	T	T	A	A	C	G	G	G	A	G	G	T	G	C	C	T	T	C	C	T	G	G	G	C	A	A	C	A	G	G	A	G	C	C	A	C	T	G	G	C	A	G	G	A	A	A	G	T	G	C	A	G	C	353				

RESULT	9
AF200708	
LOCUS	
DEFINITION	
ACCSSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	

AUTHORS	Conklin, D.S. and Beach, D.H.
TITLE	CCBR1, novel CD98 light chain implicated in redox control and calcium signaling
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1874)
AUTHORS	Conklin, D.S. and Beach, D.H.
TITLE	Direct Submission
JOURNAL	Submitted (01-NOV-1999) Cold Spring Harbor Labs, 1 Bungtown Rd, Cold Spring Harbor, NY 11724, USA
FEATURES	Location/Qualifiers

source

```

CDS
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="fibrosarcoma HT1080"
246..1751
/note="X-c transporter; CD98 light chain"
/codon_start=1
/product="calcium channel blocker resista
CCBR1"

```

CDS

```

430: 1.1751
/note="X-c transporter; CD9
/codon_start=1
/product="calcium channel p
CCBR1"
/protein_id="AAG35592.1"
/db_xref="GI:11493652"

```

BASE COUNT	490 a	407 c	422 g	555 t
ORIGIN	GIGFVITLTGVPAYYLFLLWDKKPRWF			

Alignment Scores:

Pred. No.:	6,11e-17	Length:	1874
Score:	221.00	Matches:	47
Percent Similarity:	70.00%	Conservative:	9
Best Local Similarity:	58.75%	Mismatches:	18
Query Match:	41.54%	Indels:	6
DB:	9	Gaps:	2

US-09-854-133-586 (1-97) x AF200708 (1-1874)

19	GlnThrGluLeuArgLysLysGluArgLys---	LysLysArgGluArgLysPheGlnAla	37
	:	:	
139	CAGTCTGAAGCAGAGGAAGATCGATCAGTACACCCAGAGACACCAAGTTGAAAGT	198	
38	AsnCysGlyIleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHis	57	
	::::    :	:	
199	TTTGTTTTCTTTCCCTCTGTATTATTTTCCCCCGTGTCCTCTACTATGGTCA-----	252	

58	TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet	77
253	-----GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACCTGCAGGAAATG	303
78	LeuThrGlyGlyCysLeuProTyrAlaThrArgSerHisLeuGlyArgArgLysCysSer	97
304	TTAACGGGAGGCTGCCTTCCCTGGGCAACAGGAGCCACCTGGGCAGGAGAAAGTGCAGC	363

RESULT 10  
AB040875 LOCUS  
DEFINITION  
AB040875 Homo sapiens hxCT mRNA for cystine/glutamate exchanger, complete cds.  
ACCESSION  
AB040875  
VERSION  
AB040875.1 GI:13516845  
KEYWORDS  
SOURCE  
Homo sapiens adult cDNA to mRNA  
cysteine  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913

ORGANISM

REFERENCE

1 (sites)

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae; Homo.

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Craniata; Vertebrata; Euteleostomi





US-09-854-133-586 (1-97) x BC012087 (1-2155)

QY	19	GlnThrGluLeuArgLysLysGluArgLys---	LysLysArgGluArgLysPheGlnAla	37
		:	:	
Db	54	CAGTCTGAAGCAGAGGACATCGATCAGTAACACCAAGAGACACCAAGTTGAAAGT	113	
QY	38	AsnCysGlyIleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHis	57	
		:    :	:    :	
Db	114	TTTGTTTCTTCCCTCTGTTTTATTTTCCCCCGTGTGTCTCCCTACTATGGTCA	167	
QY	58	TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet	77	
		:    :	:    :	
Db	168	-----GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACCTGCAGGGGAAATG	218	
QY	78	LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer	97	
		:    :	:    :	
Db	219	TTAACGGGAGGCTGCCCTTCCCTGGGCAACAAGGAGGCCACCTGGCGAGGAGAAAGTGCAGC	278	

Alignment Scores:		
Pred. No.:	8.41e-17	2482
Score:	221.00	47
Percent Similarity:	70.00%	9
Best Local Similarity:	58.75%	18
Query Match:	41.54%	6
DB:	9	2
	Length:	
	Matches:	
	Conservative:	
	Mismatches:	
	Indels:	
	Gaps:	

US-09-854-133-586 (1-97) x AF252872 (1-2482)

QY 19 GlnThrGluLeuArgLysLysGluArgLys--LysLysArgGluArgLysPheGlnAla 37

Db	125	:		:::		:::		:	184
		CAC	TCTGA	AGCAGAG	ACATCG	ATCAGTAA	CAACACCA	AGAGACACCA	AGTTGAAAGT
QY	38	AsnCysGly	IleAsp	PheIleIle	PheTrp	IlePheTrp	IleLeuLeu	PheSerHis	His 57
Db	185	TTTGT	TTTCTTT	CCCTCTG	TTTTAT	TTTCCCC	CGTGTGT	CCCTACTAT	GGTCA ----- 238
QY	58	TrpIleGln	GluSerLeu	CysPro	ProSerPro	LysGluVal	ThrCysArg	GluMet	77
Db	239	-----	GAA	GCCTG	TGTGCC	ACCATCT	CCAAAGG	AGGTTACCT	GCAAGGAAATG 289
QY	78	LeuThrGly	GlyCysLeu	ProTrpAla	ThrArgSerHis	LeuGlyArg	ArgLysCys	Ser 97	
Db	290	TTA	ACGGG	AGGCTGC	CTTCCCT	GGGCAACA	AGGAGCC	ACCTGGG	CAGGAGAAATGCAGC 349
RESULT	13								
HSA277882									
LOCUS		HSA277882			3144 bp	mRNA		linear	PRI 04-JAN
DEFINITION		Homo sapiens mRNA for cystine/glutamate transporter (XCT gene)							
ACCESSION		AJ277882							
VERSION		AJ277882.1		GI:18073361					
KEYWORDS		cystine/glutamate transporter; XCT gene.							
SOURCE		human.							

[illegible]



Search completed: May 11, 2003, 14:59:28  
Job time : 2223.24 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run On: May 11, 2003, 12:41:29 ; Search time 210.31 Seconds  
(without alignments)  
1038.677 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHASLGDSSETLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delopt 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_173955\_327/app\_query.fasta\_1.462  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptp -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn\_1.1\_275\_@runat\_05052003\_173955\_327 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	532	100.0	337	23	AAD23462	Human lung tumour-
2	527	99.1	5981	23	AAD23461	Human lung tumour-
3	513	96.4	2239	23	AAD23460	Human lung tumour-
4	221	41.5	575	24	ABL37718	Human colon tumour
5	221	41.5	596	24	ABK27651	Human colon cancer
6	221	41.5	800	20	AAZ16609	Human gene express
7	211.5	39.8	1958	22	ABA09201	Human cystine/Glu
8	211.5	39.8	1958	22	AAH98573	Human EST-derived
9	206.5	38.8	123	22	AAK68639	Human immune/haema
10	104.5	19.6	18408	22	AAL04753	Human reproductive
11	104.5	19.6	18408	22	AAL04754	Human reproductive
12	104.5	19.6	18408	23	ABL97656	Human testicular a
13	104.5	19.6	18408	23	ABL97657	Human testicular a
14	102	19.2	971	21	AAC58883	Human tumour suppr
15	100	18.8	3424	22	AAL06199	Human reproductive
16	100	18.8	3424	23	ABL98764	Human testicular a
17	99	18.6	109906	24	ABK94411	DNA encoding endot
18	98.5	18.5	4013	22	AAK68715	Human immune/haema
19	98.5	18.5	23164	22	AAK79678	Human immune/haema
20	97.5	18.3	154465	24	AAD28763	Human AKAP allelic
21	97.5	18.3	158245	24	AAD28762	Human AKAP allelic
22	97.5	18.3	161425	22	AAH02340	Human AKAP10 gene
23	97.5	18.3	162025	22	AAH02339	Human AKAP10 gene
24	97.5	18.3	162025	24	AAD28758	Human AKAP allelic
25	97.5	18.3	162025	24	AAD28759	Human AKAP allelic
26	97	18.2	178896	24	ABQ88146	Human osteoblast d
27	96.5	18.1	9358	24	ABK84319	Human cDNA differe
28	96.5	18.1	9358	24	ABL61905	Colon adenocarcino
29	96	18.0	32249	22	ABA17155	Human nervous syst
30	96	18.0	107820	22	AAD16230	Human ATP-binding
31	95.5	18.0	6250	22	AAK79549	Human immune/haema
32	95	17.9	752	22	AAS36489	Human cardiovascular
33	95	17.9	2138	22	AAL06186	Human reproductive
34	95	17.9	2138	23	ABL98751	Human testicular a
35	94.5	17.8	160552	22	AAD02697	Human glycosyl sul
36	94	17.7	2824	21	AAC59843	Human secreted pro
37	94	17.7	38771	22	AAK81036	Human immune/haema
38	93.5	17.6	18132	22	AAK67033	Human immune/haema
39	93.5	17.6	18132	22	AAK79622	Human immune/haema
40	93.5	17.6	47319	22	AAK64813	Human immune/haema
41	93.5	17.6	47319	22	AAK72230	Human immune/haema
42	93.5	17.6	236303	22	AAS11614	Human genomic DNA
43	93	17.5	2591	22	AAH18566	Human cDNA sequenc
44	93	17.5	28690	22	AAF57718	Human IL4Ralpha ge
45	93	17.5	31813	22	ABA08137	Human ovarian and

ALIGNMENTS

RESULT 1  
AAD23462  
ID AAD23462 standard; cDNA; 337 BP.  
XX  
AC AAD23462;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE Human lung tumour-specific 20E10 5' cDNA.  
XX  
KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
XX antisense-therapy; vaccine; immune response; lung cancer; 20E10; ss.  
OS Homo sapiens.  
XX  
PN WO200172295-A2.  
XX  
PD 04-OCT-2001.

```
XX 28-MAR-2001; 2001WO-US09991.
XX 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX (CORI-) CORIXA CORP.
XX
PI Reed SG, Lodes MJ, Mohamath R, Secrlist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX
DR WPI; 2001-639201/73.
XX
PT New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer -
XX
PS Claim 1; Page 334; 378pp; English.
XX
XX The invention relates to isolated lung tumour-specific proteins and
CC their corresponding cDNA molecules. Lung tumour-specific proteins and
CC their antigen-presenting cells are useful for stimulating and/or
CC expanding T cells specific for a tumour protein, and for inhibiting
CC the development of cancer. The invention also relates to a composition
CC useful for stimulating an immune response, and for treating cancer. The
CC lung tumour specific oligonucleotide is useful in gene therapy and for
CC diagnosis, detection and treatment of lung cancer. The present sequence
CC is a cDNA encoding human lung tumour-specific protein.
XX
SQ Sequence 337 BP; 103 A; 60 C; 93 G; 81 T; 0 other;

Alignment Scores:
Pred. No.: 5.82e-54 Length: 337
Score: 532.00 Matches: 97
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0

US-09-854-133-586 (1-97) x AAD23462 (1-337)

QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20
Db 5 GAGGTTGAAGTGAGCAGAGATCATGCCAGCCTGGTGACAGTGAGACTCTGTCTCAAACA 64
QY 21 GluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 40
Db 65 GAATTAAGGAAAAAGAAAGAAAGAAAGAGAGAGAGAGAAATTCACGGCCAAATGTGGC 124
QY 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGln 60
Db 125 ATAGATTTTATCATATTTCTGGATTTTGGATTCTTTTGTTCATCACTGGATTGAG 184
QY 61 GluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGly 80
Db 185 GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACCTGCAGGGAATGTTAACGGGA 244
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97
Db 245 GGCTGCCTTCCCTGGGCAACAAGGAGGCCACCTGGGCAGGAGAGAAAGTGCAGC 295

RESULT 2
AAD23461
ID AAD23461 standard; cDNA; 5981 BP.
XX
AC AAD23461;
XX
DT 26-FEB-2002 (first entry)
XX
DE Human lung tumour-specific 14F10 full length cDNA.
```

```
XX Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW antisense-therapy; vaccine; immune response; lung cancer; 14F10; ss.
XX Homo sapiens.
XX WO200172295-A2.
XX 04-OCT-2001.
XX
XX 28-MAR-2001; 2001WO-US09991.
XX
XX 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX (CORI-) CORIXA CORP.
XX
PI Reed SG, Lodes MJ, Mohamath R, Secrlist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX
DR WPI; 2001-639201/73.
XX
PT New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer -
XX
PS Claim 1; Page 332-334; 378pp; English.
XX
XX The invention relates to isolated lung tumour-specific proteins and
CC their corresponding cDNA molecules. Lung tumour-specific proteins and
CC their antigen-presenting cells are useful for stimulating and/or
CC expanding T cells specific for a tumour protein, and for inhibiting
CC the development of cancer. The invention also relates to a composition
CC useful for stimulating an immune response, and for treating cancer. The
CC lung tumour specific oligonucleotide is useful in gene therapy and for
CC diagnosis, detection and treatment of lung cancer. The present sequence
CC is a cDNA encoding human lung tumour-specific protein.
XX
SQ Sequence 5981 BP; 1858 A; 1029 C; 1098 G; 1996 T; 0 other;

Alignment Scores:
Pred. No.: 1.18e-51 Length: 5981
Score: 527.00 Matches: 96
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.06% Indels: 0
DB: 23 Gaps: 0

US-09-854-133-586 (1-97) x AAD23461 (1-5981)

QY 2 ValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThrGlu 21
Db 3 GTTGAAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTGAGACTCTGTCTCAAACAGAA 62
QY 22 LeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGlyIle 41
Db 63 TTAAGGAAAAAGAAAGAAAGAAAGAGAGAGAGAGAAATTCACGGCCAATGTGGCATA 122
QY 42 AspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGlnGlu 61
Db 123 GATTTTATCATATTTCTGGATTTTGGATTCTTTTGTTCATCACTGGATTGAGAA 182
QY 62 SerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGlyGly 81
Db 183 AGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACCTGCAGGGAATGTTAACGGGAGGC 242
QY 82 CysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97
Db 243 TGCCTTCCCTGGGCAACAAGGAGGCCACCTGGGCAGGAGAAAGTGCAGC 290
```

```
RESULT 3
AAD23460
ID AAD23460 standard; cDNA; 2239 BP.
XX
AC AAD23460;
XX
DT 26-FEB-2002 (first entry)
XX
DE Human lung tumour-specific 19A4 cDNA.
XX
KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW antisense-therapy; vaccine; immune response; lung cancer; 19A4; ss.
XX
OS Homo sapiens.
XX
PN WO200172295-A2.
XX
PD 04-OCT-2001.
XX
PF 28-MAR-2001; 2001WO-US09991.
XX
PR 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Lodes MJ, Mohamath R, Secríst H, Benson DR, Indirías CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX
DR WPI; 2001-639201/73.
XX
PT New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer -
XX
PS Claim 1; Page 332; 378pp; English.
XX
CC The invention relates to isolated lung tumour-specific proteins and
CC their corresponding cDNA molecules. Lung tumour-specific proteins and
CC their antigen-presenting cells are useful for stimulating and/or
CC expanding T cells specific for a tumour protein, and for inhibiting
CC the development of cancer. The invention also relates to a composition
CC useful for stimulating an immune response, and for treating cancer. The
CC lung tumour specific oligonucleotide is useful in gene therapy and for
CC diagnosis, detection and treatment of lung cancer. The present sequence
CC is a cDNA encoding human lung tumour-specific protein.
XX
SQ Sequence 2239 BP; 619 A; 444 C; 493 G; 683 T; 0 other;

Alignment Scores:
Pred. No.: 1.39e-50 Length: 2239
Score: 513.00 Matches: 97
Percent Similarity: 84.35% Conservative: 0
Best Local Similarity: 84.35% Mismatches: 0
Query Match: 96.43% Indels: 18
DB: 23 Gaps: 1

US-09-854-133-586 (1-97) x AAD23460 (1-2239)

QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20
Db 2 GAGGTTGAAGTGACAGAGATCATGCCAGCCTGGGTGACAGTGAGACTCTGTCTCAAACA 61
QY 21 GluLeuArgLysLysGluArgLysLysLysLysLysLysPheGlnAlaAsnCysGly 40
Db 62 GAATTAAGGAAAAAGAAAGAAAGAAAGAGAGAGAGAGAAATCCAGGCCAATTGTGGC 121
QY 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHistrpIleGln 60
Db 122 ATAGATTTTATCATATCTGGATTTTGGATTTCTTTGTTCTTCTCATCATGGATTTCAG 181

61 GluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGly 80
182 GAAAGCCTGTTGTCTCCACCATCTCCAAAGGAGGTACCTGCAGGGAAATGTTAACGGGA 241
81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArg----- 93
242 GGCTGCCTTCCCTGGCAACAAGAGAGACCCACCTGGGACAGGAGCGCTTTTCAGGAAGAGACG 301
94 -----ArgLysCysSer 97
302 CCTTTTCAGGAAGAGAGCGCTTTTCAGGAAGAGAGAGAAAGTGCGAGC 346

RESULT 4
ABL37718
ID ABL37718 standard; cDNA; 575 BP.
XX
AC ABL37718;
XX
DT 08-APR-2002 (first entry)
XX
DE Human colon tumour antigen polynucleotide SEQ ID NO:1307.
XX
KW Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
KW colon tumour metastatic antigen; diagnosis; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200196388-A2.
XX
PD 20-DEC-2001.
XX
PF 08-JUN-2001; 2001WO-US18557.
XX
PR 09-JUN-2000; 2000US-210899P.
PR 20-FEB-2001; 2001US-270216P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Jiang Y, Harlocker SL, Secríst H;
XX
DR WPI; 2002-114514/15.
XX
PT Novel isolated colon tumor polynucleotide differentially expressed in
PT colon tumor or colon metastatic tumor and polypeptides encoded by them,
PT useful for inhibiting development of cancer in patient -
XX
PS Claim 1; SEQ ID 1307; 105pp; English.
XX
CC ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I)
CC which were isolated from human colon tumour and colon metastatic tumour
CC cDNA libraries. (I) have cytostatic activity and can be used in vaccine
CC production. (I) can be used for stimulating and/or expanding T cells
CC specific for a tumour protein on contact with the T cells. They are also
CC useful for inhibiting the development of cancer in a patient. (I) can be
CC used as probes or primers for nucleic acid hybridisation, for preparing
CC mutant species primers, or primers for use in genetic constructions. (I)
CC can be used in the diagnosis of a colon tumour.
XX
SQ Sequence 575 BP; 151 A; 122 C; 153 G; 141 T; 8 other;

Alignment Scores:
Pred. No.: 8.13e-17 Length: 575
Score: 221.00 Matches: 47
Percent Similarity: 70.00% Conservative: 9
Best Local Similarity: 58.75% Mismatches: 18
Query Match: 41.54% Indels: 6
DB: 24 Gaps: 2

US-09-854-133-586 (1-97) x ABL37718 (1-575)

QY 19 GlnThrGluLeuArgLysLysLysGluArgLys---LysLysArgGluArgLysPheGlnAla 37
||||:|||| ||| ::| ||| |||:| ||| :|:|:|
```

Db 142 CAGTCTGAAAGCAGAGGAGACATCGATCAGTAACACCAAGAGACACCAAGTTGAAAGT 201  
QY 38 AsnCysGlyIleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHis 57  
Db 202 TTTGTTTCTTTCCCTCTGTTTATTTTCCCTGTTGTTCCCTACTATGTTCA----- 255  
QY 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77  
Db 256 -----GAAAGCCTGTTGTCTCCACCATCTCCAAAGGAGGTTACCTGCAGGGAATG 306  
QY 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
Db 307 TTAACGGGAGGCTGCCTTCCCTGGCAACAAGGAGGCCACTGGGCAGGAGAAAGTGCAGC 366

## RESULT 5

ABK27651/c

ID ABK27651 standard; cDNA; 596 BP.

XX AC ABK27651;

XX AC ABK27651;

DT 09-APR-2002 (first entry)

XX Human colon cancer expressed sequence tag, Seq ID no 88.

KW Human; colon cancer; T cell expansion; tumour; EST; gene; ss;

KW expressed sequence tag.

XX Homo sapiens.

OS Homo sapiens.

XX Homo sapiens.

PN WO200196390-A2.

XX WO200196390-A2.

XX 20-DEC-2001.

XX 08-JUN-2001; 2001WO-US18577.

XX 09-JUN-2000; 2000US-210821P.

PR 18-DEC-2000; 2000US-256571P.

PR 10-MAY-2001; 2001US-290240P.

XX 10-MAY-2001; 2001US-290240P.

XX (CORI-) CORIXA CORP.

PA (CORI-) CORIXA CORP.

XX Jiang Y, Hepler WT, Clapper JD, Wang A, Secrist H;

PI Jiang Y, Hepler WT, Clapper JD, Wang A, Secrist H;

XX WPI; 2002-139708/18.

DR WPI; 2002-139708/18.

XX Novel isolated polynucleotide encoding a polypeptide comprising a

PT portion of colon tumour protein, useful for detection, diagnosis and

PT therapy of human colon cancer

XX Claim 1; Page 174; 220pp; English.

PS Claim 1; Page 174; 220pp; English.

XX The invention relates to an isolated polynucleotide (I) encoding a

CC polypeptide (II) comprising at least a portion of a colon tumour

CC protein. (I), (II) and antibody (III) to (II) are useful for determining

CC the presence of a cancer in a patient. (I), (II) or antigen presenting

CC cells expressing (I) is useful for stimulating and/or expanding T cells

CC specific for a tumour protein, by contacting T cells with (I), (II) or

CC antigen-presenting cells that express (I), under conditions and for a

CC time sufficient to permit the stimulation and/or expansion of T cells.

CC (I), (II), or antigen presenting cells that express (II) are useful for

CC treating colon cancer in a patient by incubating CD4+ and/or CD8+ T cells

CC isolated from a patient with (I), (II) or antigen presenting cells that

CC express (II), such that T cells proliferate, and administering to the

CC patient an effective amount of the proliferated T cells, thus inhibiting

CC the development of a cancer in the patient. (I) or (II) is useful in

CC vaccines and pharmaceutical compositions for prevention and treatment

CC of colon malignancies and for the diagnosis and monitoring of such

CC cancers. (I), (II) or (III) is useful for detection, diagnosis and/or

CC therapy of human colon cancer. (I) is useful as a probe or primer for

CC nucleic acid hybridisation, and in the design and preparation of

CC ribozyme molecules for inhibiting expression of (II) in tumour cells.

CC ABK27564-ABK27807 represent novel human colon cancer coding

CC sequences and primers of the invention.

XX

SQ Sequence 596 BP; 160 A; 151 C; 126 G; 158 T; 1 other;

## Alignment Scores:

Pred. No.: 8.54e-17 Length: 596  
Score: 221.00 Matches: 47  
Percent Similarity: 70.00% Conservative: 9  
Best Local Similarity: 58.75% Mismatches: 18  
Query Match: 41.54% Indels: 6  
DB: 24 Gaps: 2

US-09-854-133-586 (1-97) x ABK27651 (1-596)

QY

19

GlnThrGluLeuArgLysLysGluArgLys---

LysLysArgGluArgLysPheGlnAla 37

Db

481

CAGTCTGAAAGCAGAGGAGACATCGATCAGTAACACCAAGAGACACCAAGTTGAAAGT 422

QY

38

AsnCysGlyIleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHis 57

Db

421

TTTGTCTTTCTTTCCCTCTGTTTATTTTCCCTGTTGTTCCCTACTATGTTCA----- 368

QY

58

TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77

Db

367

-----GAAAGCCTGTTGTGTCTCCACCATCTCCAAAGGAGGTTACCTGCAGGGAATG 317

QY

78

LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97

Db

316

TTAACGGGAGGCTGCCTTCCCTGGCAACAAGGAGGCCACTGGGCAGGAGAAAGTGCAGC 257

RESULT 6

AAZ16609

ID AAZ16609 standard; cDNA; 800 BP.

XX AAZ16609;

XX AAZ16609;

DT 12-OCT-1999 (first entry)

XX Human gene expression product cDNA sequence SEQ ID NO:4079.

DE Human; gene; gene expression product; diagnosis; therapy; probe;

KW detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX Homo sapiens.

OS Homo sapiens.

XX WO9938972-A2.

PN WO9938972-A2.

XX 05-AUG-1999.

PD 05-AUG-1999.

XX 28-JAN-1999; 99WO-US01619.

PF 28-JAN-1999; 99WO-US01619.

XX 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

XX (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

DR Novel human genes and their expression products which are

XX differentially expressed in different cell types

PT Claim 1; Page 1934; 2479pp; English.

XX Claim 1; Page 1934; 2479pp; English.

XX





US-09-854-133-586 (1-97) x ABA09201 (1-1958)

QY 38 AsnCysGlyIleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisHis 57  
Db 2 AATTCGGGTGCGACGATTTCGTTTCCCTCTGTATTATTTTCCCGGTGTCCCTAC 61  
QY 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77  
Db 62 TATGG-TCAGAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTACCTGCAGGAAATG 120  
QY 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
Db 121 TTAACGGGAGGCTGCCTTCCCTGGGCAACAAGGAGCCACCTGGGAGGAGAAAGTGCAGC 180

RESULT 8

AAH98573  
ID AAH98573 standard; cDNA; 1958 BP.

XX AAH98573;

AC 12-OCT-2001 (first entry)

DT Human EST-derived coding sequence SEQ ID NO: 430.

XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder;  
KW biodiversity; gene therapy; nutrition; ss.

XX Homo sapiens.

OS WO200154477-A2.

PN 02-AUG-2001.

PD 25-JAN-2001; 2001WO-US02687.

XX 25-JAN-2000; 2000US-0491404.

PR 17-JUL-2000; 2000US-0617746.

PR 03-AUG-2000; 2000US-0631451.

PR 15-SEP-2000; 2000US-0663870.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

PI Cao Y, Drmanac RA, Zhang J, Werhman T;

XX WPI; 2001-476164/51.

DR P-PSDB; AAM23914.

XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
PT antibodies and research use -  
PS Claim 1; Page 484; 1275pp; English.

CC The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a cDNA  
XX of the invention.

SQ Alignment Scores:

Pred. No.: 5.83e-15 Length: 1958  
Score: 211.50 Matches: 41  
Percent Similarity: 78.33% Conservative: 6  
Best Local Similarity: 68.33% Mismatches: 13  
Query Match: 39.76% Indels: 1  
DB: 22 Gaps: 0

US-09-854-133-586 (1-97) x AAH98573 (1-1958)

QY 38 AsnCysGlyIleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisHis 57  
Db 2 AATTCGGGTGCGACGATTTCGTTTCCCTCTGTATTATTTTCCCGGTGTCCCTAC 61  
QY 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77  
Db 62 TATGG-TCAGAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTACCTGCAGGAAATG 120  
QY 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
Db 121 TTAACGGGAGGCTGCCTTCCCTGGGCAACAAGGAGCCACCTGGGAGGAGAAAGTGCAGC 180

RESULT 9

AAK68639

ID AAK68639 standard; DNA; 123 BP.

XX AAK68639;

AC AAK68639;

XX 06-NOV-2001 (first entry)

DT Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23451.  
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

XX WO200157182-A2.

PN 09-AUG-2001.

PD 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 30-JUN-2000; 2000US-0214886.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-483426/52.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -

XX Disclosure; SEQ ID NO 23451; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.

XX SQ Sequence 123 BP; 32 A; 31 C; 37 G; 23 T; 0 other;

Alignment Scores:

Pred. No.:	5.1e-16	Length:	123
Score:	206.50	Matches:	37
Percent Similarity:	94.87%	Conservative:	0
Best Local Similarity:	94.87%	Mismatches:	1
Query Match:	38.82%	Indels:	1
DB:	22	Gaps:	1

US-09-854-133-586 (1-97) x AAK686639 (1-123)

QY 58 TriPleGlnGluSerLeuLeuCysProCysProLysGluValThrCysArgGluMet 77  
||| |||||||||||||||||||||||||||||||||||||||||||||

Db 8 TGG---TCAGAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTACCTGCAGGAAATG 64







PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI; 2001-465570/50.  
XX  
PT Isolated nucleic acid molecule encoding a reproductive system antigen -  
PT is used in preventing, treating or ameliorating a medical condition  
XX  
PS Disclosure; SEQ ID NO 7442; 1297pp + Sequence Listing; English.  
XX  
CC The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a genomic sequence encoding a  
CC protein of the invention.  
XX  
SQ Sequence 18408 BP; 4392 A; 4705 C; 4957 G; 4354 T; 0 other;

Alignment Scores:  
Pred. No.: 0.591 Length: 18408  
Score: 104.50 Matches: 32  
Percent Similarity: 53.93% Conservative: 16  
Best Local Similarity: 35.96% Mismatches: 23  
Query Match: 19.64% Indels: 18  
DB: 22 Gaps: 4

US-09-854-133-586 (1-97) x AAL04754 (1-18408)

QY 2 ValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThrGlu 21  
Db 15928 GTGGAGATCGCACCACTGCACCTCAGCCTGGGTGATAGCAAACACTGTCTCAAAA --- 15984  
QY 22 LeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGlyIle 41  
Db 15985 ---AAGAAAGAAGAGAAGAAAAAATAATCACCCCTTTATT-----CTT 16026  
QY 42 AspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHis-----His 57  
Db 16027 TTTTTTTTGCTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTTCACCTCGTGGCCAGGC 16086  
QY 58 Trp-IleGlnGluSerLeuLeuCys-----ProProSerProLysGI 71  
Db 16087 TGGAGTGCAATGGCCGATCTCTGCTCACTACAACAACCTCCACCTCCCGGGTTCAAGCG 16146  
QY 71 uValThrCysArgGluMetLeuThr 79  
Db 16147 ATTCTCTGCCCTCAGCCTCCTGAGT 16171

RESULT 12  
ABL97656  
ID ABL97656 standard; DNA; 18408 BP.  
XX  
AC ABL97656;  
XX  
DT 21-JUN-2002 (first entry)  
XX  
DE Human testicular antigen encoding DNA fragment SEQ ID NO: 2308.  
XX  
KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;  
KW reproductive system disorder; urinary system disorder; gene therapy;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
KW gastrointestinal disease; infection; cytostatic; gene; ds.  
OS Homo sapiens.  
XX  
PN WO200155317-A2.  
XX  
PD 02-AUG-2001.

XX  
PF 17-JAN-2001; 2001WO-US01329.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483232/52.  
XX

[illegible]

RESULT 13	
ABL97657	
ID	ABL97657 standard; DNA; 18408 BP.
XX	
AC	ABL97657;
XX	
DT	21-JUN-2002 (first entry)
XX	
DE	Human testicular antigen encoding DNA fragment SEQ ID NO: 2309.
XX	
KW	Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW	reproductive system disorder; urinary system disorder; gene therapy;
KW	cardiovascular disorder; respiratory disorder; neurological disorder;
KW	gastrointestinal disease; infection; cytostatic; gene; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200155317-A2.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01329.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.



PR	13-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.



PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX		
XX		
PA	(HUMA-)	HUMAN GENOME SCI INC.
XX		
PI	Rosen CA,	Barash SC, Ruben SM;
XX		
XX	WPI;	2001-465570/50.
DR		
XX		
PT	Isolated nucleic acid molecule encoding a reproductive system antigen	-
PT	is used in preventing, treating or ameliorating a medical condition	
XX		
PS	Disclosure; SEQ ID NO 8887; 1297pp + Sequence Listing; English.	
XX		
CC	The present invention provides the protein and coding sequences of a	
CC	number of human reproductive system related antigens. These can be used	
CC	in the prevention and treatment of reproductive system disorders,	
CC	including cancer. The present sequence is a genomic sequence encoding a	
CC	protein of the invention.	
XX		
SQ	Sequence 3424 BP; 676 A; 1024 C; 1003 G; 721 T; 0 other;	

Alignment Scores:			
Pred. No.:	0.2	Length:	3424
Score:	100.00	Matches:	30
Percent Similarity:	50.00%	Conservative:	12
Best Local Similarity:	35.71%	Mismatches:	26
Query Match:	18.80%	Indels:	16
DB:	22	Gaps:	3
US-09-854-133-586 (1-97) x AAL06199 (1-3424)			
QY	3	GluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThrGluLeu	22
Db	798	GAGATCGCGCCACTGCACCTCCAGCCTGGGCGAC-----AGACTG	760
QY	23	ArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGlyIleAsp	42
Db	759	AGACTCCGTATCAGAAAAAAGAAAAAGAACTCACTGGGAAGTTCCCGCTTCGCGG	700
QY	43	PheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGlnGluSer	62
Db	699	ACCATTACATCCAGGGTCTGTGGCTCTTA-----GCTTCC	664
QY	63	LeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCys	82
Db	663	TTGCTGGACCTGGCAGTCCCA---ACTGTCTCCAGGAGAGAGAAAGTTTCCGCTTCGCGG	607
QY	83	LeuProTrpAla	86
Db	606	CTTTTCTGGGCA	595

Search completed: May 11, 2003, 14:16:17  
Job time : 222.31 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 14:25:24 ; Search time 53.2212 Seconds  
(without alignments)  
558.943 Million cell updates/sec

**Title:** US-09-854-133-586  
**Perfect score:** 532  
**Sequence:** 1 EVEVSRDHASLGDSSETLSQT.....LTGGCLPWATRSHLGRKKCS 97

Scoring table:	BLOSUM62	
	Xgapop	10.0 , Xgapext 0.5
	Ygapop	10.0 , Ygapext 0.5
	Fgapop	6.0 , Fgapext 7.0
	Delop	6.0 , Delext 7.0

```
Searched: 441362 segs, 153338381 residues
Total number of hits satisfying chosen parameters: 882724
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09854133/runat_05052003_173956_361/app_query.fasta_1.462
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=ini -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosom62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09854133@cGN_1_1_47 @runat_05052003_173956_361 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -XGAPEXT=0.5 -DEPOP=6 -DELEXT=7

```

```
Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	102	19.2	971	4	US-09-535-008-39	Sequence 39, Appl
C 2	91.5	17.2	99500	4	US-09-798-096-10	Sequence 10, Appl
C 3	85.5	16.1	2455	4	US-08-406-030A-4	Sequence 4, Appl
4	84.5	15.9	17327	1	US-07-906-871-15	Sequence 15, Appl
C 5	84	15.8	5529	3	US-08-869-696-1	Sequence 1, Appl
6	84	15.8	22481	4	US-08-367-841A-43	Sequence 43, Appl
7	84	15.8	22481	5	PCT-US95-07201-43	Sequence 43, Appl
8	84	15.8	22484	4	US-09-875-223-2	Sequence 2, Appl
9	83.5	15.7	3224	4	US-08-965-729A-2	Sequence 2, Appl
C 10	83.5	15.7	4895	3	US-09-053-866-1	Sequence 1, Appl
C 11	83.5	15.7	4895	4	US-09-479-130-1	Sequence 1, Appl
12	83	15.6	4129	2	US-08-370-319C-12	Sequence 12, Appl

## ALIGNMENTS

```

RESULT 1
US-09-535-008-39
; Sequence 39, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR
; TITLE OF INVENTION: AND OTHER CANCER
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/5
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 971
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-535-008-39

```

Alignment Scores:		
Pred. No.:	0.00353	Length:
Score:	102.00	Matches:
Percent Similarity:	45.92%	Conservative:
Percent Local Similarity:	33.67%	Mismatches:
Query Match:	19.17%	Indels:
DB:	4	Gaps:
		971
		33
		12
		28
		25
		6

US-09-854-133-586 (1-97) x US-09-535-008-39 (1-971)

Qy	3	GluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThrGluLeu	22
	:::	:::	::
Db	22	GAGATCAGCCCACTGCACCTCTAGCCTGGTGAC-----AGAGCGGAGACTC	66



US-09-854-133-586 (1-97) x US-08-406-030A-4 (1-2455)  
QY 3 GluValSerArgAspHisAlaSerLeuGlyAsp---SerGluThrLeuSerGlnThrGlu 21  
Db 1352 GAGATCGCACCATTCGCACTCCAGCTGGGCAATAAGAGTGAAGTCCGCTCTCAAAAAA 1293  
QY 22 LeuArgLysLysGluArgLysLysArgGluArgLysPheGlnAlaAsnCysGlyIle 41  
Db 1292 AAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1245  
QY 42 AspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGlnGlu 61  
Db 1244 TCCTTTTGTGAGGA-----CTCCTCAGTGAA-----TCGGAC 1212  
QY 62 SerLeuLeuCysProSerProSerProLysGlu-ValThrCysArgGluMetLeuThr----- 79  
Db 1211 ATGATTGGTGGCCACCTACCCCTGCTGAGAGCAGCAGCATGATATGTATACACAT 1152  
QY 80 -----GlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgAr 94  
Db 1151 ATCTTAATAGCCTTAATTTCTGAGCAACCTTGGTTG-----TTAGGTACAGAG 1104  
QY 94 gLys 95  
Db 1103 AAAA 1100

RESULT 4  
US-07-906-871-15  
; Sequence 15, Application US/07906871  
; Patent No. 5340739  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; APPLICANT: Avraham, Shalom  
; TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC  
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES  
; TITLE OF INVENTION: THEREOF  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1225 Connecticut Avenue, N.W., Suite 300  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/906,871  
; FILING DATE: 19920103  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/816,289  
; FILING DATE: 03 JAN 1992  
; APPLICATION NUMBER: US/07/635,544  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US89/03051  
; FILING DATE: 13-JUL-1989  
; APPLICATION NUMBER: US/07/224,035  
; FILING DATE: 13-JUL-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cimbala, Michele A  
; REGISTRATION NUMBER: 33,851  
; REFERENCE/DOCKET NUMBER: 0627.2830004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)833-7533  
; TELEFAX: (202)833-8716  
; INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:  
LENGTH: 17327 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: exon  
LOCATION: 621..753  
FEATURE:  
NAME/KEY: intron  
LOCATION: 754..9596  
FEATURE:  
NAME/KEY: exon  
LOCATION: 9597..9744  
FEATURE:  
NAME/KEY: intron  
LOCATION: 9745..16396  
FEATURE:  
NAME/KEY: exon  
LOCATION: 16397..17327  
US-07-906-871-15  
Alignment Scores:  
pred. No.: 23.5 Length: 17327  
Score: 84.50 Matches: 24  
Percent Similarity: 54.67% Conservative: 17  
Best Local Similarity: 32.00% Mismatches: 11  
Query Match: 15.88% Indels: 23  
DB: 1 Gaps: 5  
US-09-854-133-586 (1-97) x US-07-906-871-15 (1-17327)  
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAsp-----SerGluThr 16  
Db 4329 GAGGTTGCACTGAGCCCAAGATCATGCCACTGCACCTGCCAGCTGGCGTAGACTGAGACT 4388  
QY 17 LeuSerGlnThrGluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGln 36  
Db 4389 CTGTCTCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAATTA--- 4445  
QY 37 AlaAsnCys-----GlyIleAspPhe 43  
Db 4446 GTATATTGTGATTATGTTGAGGAAAGTTAGTACCATAATATAAAGGTATGGAC--- 4502  
QY 44 IleIlePheTrp-----IlePheTrpIleLeuLeuPheSerHis 56  
Db 4503 -----TATTGGAGAAAGTTGTTTGTCTTGGTAACATTACTCAT 4541

RESULT 5  
US-08-869-696-1/c  
; Sequence 1, Application US/08869696C  
; Patent No. 6031155  
; GENERAL INFORMATION:  
; APPLICANT: Cameron-Mills, Verena  
; APPLICANT: Lok, Finn  
; APPLICANT: Sinjorgo, Catharina Maria Cornelia  
; APPLICANT: Van Den Dool, Ronald Tako Marinus  
; APPLICANT: Caspers, Martinus Petrus Maria  
; APPLICANT: Van Zeijl-Van Der Valk, Maria Joanna  
; TITLE OF INVENTION: ARABINOXYLAN DEGRADATION  
; FILE REFERENCE: 11225.01US01  
; CURRENT APPLICATION NUMBER: US/08/869,696C  
; CURRENT FILING DATE: 1997-06-05  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 5529  
; TYPE: DNA  
; ORGANISM: barley  
US-08-869-696-1  
Alignment Scores:





Db 13335 GAGATGGGGCCTCACCATGTTGCTCAGGTTGGTCTGAAACTCCTGAGCTCAAGTGATCCA 13394

QY 70 EYSCGCGGCTGCGAA 13412

Db 13335 GAGATGGGGCCTCACCATGTTGCTCAGGTTGGTCTGAAACTCCTGAGCTCAAGTGATCCA 13394

```

Db 1123 TTCTACTTTTCCTGAGGCCAGTGCAGGAAACAAAGTTCCCTATCCTTGTTCCTCAACTAGA 1182
      |||||
      :::|||||:::
      |||||
Qy 69 Pro 69
      |||
Db 1183 CCA 1185

RESULT 10
US-09-053-866-1/c
; Sequence 1, Application US/09053866
; Patent No. 6111075
; GENERAL INFORMATION:
; APPLICANT: Xu, Wenfeng
; APPLICANT: Presnell, Scott R.
; APPLICANT: Yee, David P.
; APPLICANT: Foster, Donald C.
; TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR
; TITLE OF INVENTION: PAR4 (ZCHEMR2)
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: zymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/053,866
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Leith, Debra K
; REGISTRATION NUMBER: 32,619
; REFERENCE/DOCKET NUMBER: 98-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6674
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4895 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 176...1330
; OTHER INFORMATION:
US-09-053-866-1

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:
      5.4      4895
      83.50    35
      41.74%   13
      30.43%   37
      15.70%   31
      3        6

US-09-854-133-586 (1-97) x US-09-053-866-1 (1-4895)

Qy 1 GluValGluValSerArgAspHisAlaSerLeu-----GlyAspSerGluThr 16
      |||::: |||||
Db 4362 GAGCTTGAGTGAGCGAGATCAGCCACTGCCACTCCAGCCTGGGCAACAGCGCGAGACT 4303
      |||||

Qy 17 LeuSerGlnThrGluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGln 36

```

Db	4302	CTGTCTCAGGAA-----AAAAGAAAAAGAAAAAGAAAGAAAAAGAAAAATGGGA	4249
QY	37	AlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHis	56
Db	4248	TCACGTTTC-----TTCTGTGTTTCACGGCTCTCGGTATCCCTGAAATCAGCCAG	4198
QY	57	HisTrpIleGlnGluSerLeuLeuCysProProSerPro-----	69
Db	4197	TTCTGGGCTGAGCTCAGCGCCCGGGCCTCCTGGCCCGGTCGAGCCTGCGCCCGCCGC	4138
QY	70	-----LysGluValThrCysArgGluMetLeuThrGlyGly	81
Db	4137	TGCCCCACAGATCCCCAGCGGTGCTGACGTGCG-TGCGGTTCGGTGCTCTCCGGGAGT	4079
QY	82	CysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCys	96
Db	4078	GCGTGG--TGGGCA-----GGACGTGCGCGCTGC	4052
RESULT 11			
US-09-479-130-1/c			
; Sequence 1, Application US/09479130			
; Patent No. 6436400			
; GENERAL INFORMATION:			
; APPLICANT: Xu, Wenfeng			
; APPLICANT: Presnelli, Scott R.			
; APPLICANT: Yee, David P.			
; APPLICANT: Foster, Donald C.			
; TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR			
; TITLE OF INVENTION: PAR4 (ZCHEMR2)			
; NUMBER OF SEQUENCES: 12			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: ZymoGenetics, Inc.			
; STREET: 1201 Eastlake Avenue East			
; CITY: Seattle			
; STATE: WA			
; COUNTRY: USA			
; ZIP: 98102			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Diskette			
; COMPUTER: IBM Compatible			
; OPERATING SYSTEM: DOS			
; SOFTWARE: FastSeq for Windows Version 2.0			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/09/479,130			
; FILING DATE:			
; CLASSIFICATION:			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER:			
; FILING DATE:			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Leith, Debra K			
; REGISTRATION NUMBER: 32,619			
; REFERENCE/DOCKET NUMBER: 98-10			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: 206-442-6674			
; TELEFAX: 206-442-6678			
; TELEX:			
; INFORMATION FOR SEQ ID NO: 1:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 4895 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; FEATURE:			
; NAME/KEY: Coding Sequence			
; LOCATION: 176...1330			
; OTHER INFORMATION:			
US-09-479-130-1			
Alignment Scores:			
Pred. No.:			
Score:			
5.4			
83.50			
Length:			
Matches:			
4895			
35			

```
Percent Similarity:      41.74%      Conservative:    13
Best Local Similarity:   30.43%      Mismatches:     37
Query Match:            15.70%      Indels:         31
DB:                      4          Gaps:           6

US-09-854-133-586 (1-97) x US-09-479-130-1 (1-4895)

QY 1 GluValGluValSerArgAspHisAlaSerLeu-----GlyAspSerGluThr 16
    |||::: ||||| ||||| ||||| :::
Db 4362 GAGCTTGCA GTGAGCCGAGATCACGCCACTGCACCTCCAGCCTGGCGAACAGCGGAGACT 4303

QY 17 LeuSerGlnThrGluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGln 36
    ||||||| :::|||:::|||:::||| ||| ::::|||
Db 4302 CTGTCTCAGGAA-----AAAAAAGTAAAAGAAAAAAGATAAAGAAAAAAGAAATGGGA 4249

QY 37 AlaAsnCysGlyIleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHis 56
    ::: ||| ::: ||| ||| |||
Db 4248 TCACGTTTC-----TTCTGTGTTTCACGCGTCTCTGGGTATCCCTGAAATCAGCCAG 4198

QY 57 HisTrpIleGlnGlusLeuLeuCysProProSerPro----- 69
    ||| ||||| |||
Db 4197 TTCTGGCCTGAGCTCAGCGCCCCGGGCTCTCTGGCCCCGGTGCAGCCTGCGCCGCCGC 4138

QY 70 -----LysGluValThrCysArgGluMetLeuThrGlyGly 81
    :::||| ||| :::|||:::|||
Db 4137 TGCCCCACAGATCCCAGCGGTGCCTGACGTGCG-TGCGGTTCCGTGCTCTCCGGGAGT 4079

QY 82 CysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCys 96
    ||||| |||||:::|||
Db 4078 GCGTGG---TGGGCA-----GGACGTGGCGCTGC 4052

RESULT 12
US-08-370-319C-12
; Sequence 12, Application US/08370319C
; Patent No. 5856091
; GENERAL INFORMATION:
; APPLICANT: Brichard, Vincent; Van Pel, Aline;
; APPLICANT: Traversari, Catia; W lfel, Thomas; Coulie, Pierre;
; APPLICANT: Boon-Falleur, Thierry; De plaen, Etienne
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PERCURSOR PROCESSED TO AT
; TITLE OF INVENTION: REJECTION ANTIGEN PRESENTED BY HLA-A2
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,319C
; FILING DATE: 10-JANUARY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/272,351
; FILING DATE: 8-JULY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/032,978
; FILING DATE: 18-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, NO. 5856091man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5377.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 12:
```

```

;
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5377.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; OTHER INFORMATION: The sequence is preceded by an
; OTHER INFORMATION: unsequenced portion of from 4.7 to 5.3
; OTHER INFORMATION: kilobases
US-09-224-834-12

Alignment Scores:
Pred. No.: 4.89 Length: 4129
Score: 83.00 Matches: 28
Percent Similarity: 51.35% Conservative: 10
Best Local Similarity: 37.84% Mismatches: 12
Query Match: 15.60% Indels: 24
DB: 4 Gaps: 4

US-09-854-133-586 (1-97) x US-09-224-834-12 (1-4129)

QY 1 GluValGluValSerArgAspHisAlaSerLeu-----GlyAspSerg
||||| ||||||||| ||| ::
Db 1441 GAGGTTGCAGTGAGCGAGATCGTGCCATTATACTCCAGCTGGGCAACAGAGT
QY 17 LeuSerGlnThrGluLeuArgLysLysGluArgLysLysLysArgGluArgLysPhe
||||||| ::||| ||| ::||| ||| ::||| ||| ::||| |||
Db 1501 CTGTCTCAA-----AAAAAAAAGAAAGAAAGAAAGAAAGAAAAAAGT
QY 37 AlaAsnCysGlyIleAspPheIlePheTrpIlePheTrpIleLeuLeuPhe-
||||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 1549 -----TTATTTTACCTTCACCTTATCTCCT
QY 55 -----SerHisHisTrpIleGlnGlnSerLeuLeuCys 65
||| ::||| ||| |||||
Db 1576 GGATGTTCTCTCTTATGTAGGTACAAGGTTCTGACCTATGT 1617

RESULT 14
US-09-814-951A-3
; Sequence 3, Application US/09814951A
; Patent No. 6387661
; GENERAL INFORMATION:
; APPLICANT: SHAO, Wei et al
; TITLE OF INVENTION: ISOLATED HUMAN AMINOACYLASE, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN AMINOACYLASE
; FILE REFERENCE: CL001179
; CURRENT APPLICATION NUMBER: US/09/814,951A
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 9704
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-951A-3

Alignment Scores:
Pred. No.: 18.3 Length: 9704
Score: 82.50 Matches: 19
Percent Similarity: 74.42% Conservative: 13
Best Local Similarity: 44.19% Mismatches: 8
Query Match: 15.51% Indels: 3
DB: 4 Gaps: 2

US-09-854-133-586 (1-97) x US-09-814-951A-3 (1-9704)

QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAsp-----SerGluThrLr

```



Search completed: May 11, 2003, 16:16:03  
Job time : 85.2212 secs



GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 15:28:15 ; Search time 115.027 Seconds  
(without alignments)  
1047.953 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHASLGDSLTSLT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 783854 seqs, 621352466 residues  
Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_173958\_449/app\_query.fasta\_1.462  
-DB=Published\_Applications\_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09854133@cgn\_1.117@runat\_05052003\_173958\_449  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published\_Applications\_NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	532	100.0	337	9 US-09-854-133-442	Sequence 442, App
2	532	100.0	337	10 US-09-738-973-442	Sequence 442, App
3	527	99.1	5981	9 US-09-854-133-441	Sequence 441, App
4	527	99.1	5981	10 US-09-738-973-441	Sequence 441, App

5	513	96.4	2239	9 US-09-854-133-440	Sequence 440, App
6	513	96.4	2239	10 US-09-738-973-440	Sequence 440, App
7	221	41.5	572	9 US-10-046-935-1307	Sequence 1307, Ap
8	221	41.5	572	9 US-09-878-178-1307	Sequence 1307, Ap
9	221	41.5	572	9 US-10-146-502-1307	Sequence 1307, Ap
10	221	41.5	1861	9 US-10-163-866-30	Sequence 30, Appl
11	221	41.5	2000	9 US-10-163-866-34	Sequence 34, Appl
12	221	41.5	2482	9 US-10-163-866-29	Sequence 29, Appl
13	210.5	39.6	520	9 US-10-163-866-32	Sequence 32, Appl
14	210.5	39.6	1268	9 US-10-163-866-53	Sequence 53, Appl
15	210.5	39.6	1528	9 US-10-163-866-52	Sequence 52, Appl
16	210.5	39.6	1542	9 US-10-163-866-33	Sequence 33, Appl
17	210.5	39.6	3144	9 US-10-163-866-31	Sequence 31, Appl
18	104.5	19.6	18408	9 US-09-764-891-7441	Sequence 7441, Ap
19	104.5	19.6	18408	9 US-09-764-891-7442	Sequence 7442, Ap
20	100.5	18.9	88191	10 US-09-799-799-3	Sequence 3, Appl1
21	100	18.8	3424	9 US-09-764-891-8887	Sequence 8887, Ap
22	95	17.9	752	9 US-10-091-504-1989	Sequence 1989, Ap
23	95	17.9	752	10 US-09-764-869-1989	Sequence 1989, Ap
24	95	17.9	2138	9 US-09-764-891-8874	Sequence 8874, Ap
25	93.5	17.6	36303	9 US-10-152-724A-24	Sequence 24, Appl
26	93	17.5	28690	9 US-10-010-802-1	Sequence 1, Appl1
27	93	17.5	31813	9 US-09-764-891-9685	Sequence 9685, Ap
28	92	17.3	16086	10 US-09-764-877-2385	Sequence 2385, Ap
29	92	17.3	76798	10 US-09-880-107-3949	Sequence 3949, Ap
30	92	17.3	203654	10 US-09-820-905-3	Sequence 3, Appl1
31	91.5	17.2	32169	9 US-10-092-154-1963	Sequence 1963, Ap
32	91.5	17.2	32169	10 US-09-764-847-1963	Sequence 1963, Ap
33	91.5	17.2	465237	10 US-09-933-267A-1	Sequence 1, Appl1
34	91	17.1	3425	9 US-09-764-891-8888	Sequence 8888, Ap
35	91	17.1	10236	9 US-10-091-438-242	Sequence 242, App
36	91	17.1	10236	9 US-10-091-438-282	Sequence 282, App
37	91	17.1	10236	10 US-09-764-853-878	Sequence 878, App
38	91	17.1	52216	10 US-09-747-810-1	Sequence 1, Appl1
39	90.5	17.0	7032	9 US-10-092-154-1429	Sequence 1429, Ap
40	90.5	17.0	7032	10 US-09-764-847-1429	Sequence 1429, Ap
41	90.5	17.0	7537	9 US-10-091-504-1735	Sequence 1735, Ap
42	90.5	17.0	7537	10 US-09-764-869-1735	Sequence 1735, Ap
43	90	16.9	63000	10 US-09-780-172-18	Sequence 18, Appl
44	89.5	16.8	601	10 US-09-820-003A-22	Sequence 22, Appl
45	89.5	16.8	601	10 US-09-820-003A-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1  
US-09-854-133-442  
; Sequence 442, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodges, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 442  
; LENGTH: 337  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-854-133-442

Alignment Scores:  
Pred. No.: 2.77e-62 Length: 337  
Score: 532.00 Matches: 97  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-854-133-586 (1-97) x US-09-854-133-442 (1-337)

QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
 |||||  
 Db 5 GAGGTTGAAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTGGAGACTCTGTCTCAACA 64

QY 21 GluLeuArgLysLysGluArgLysLysArgGluValThrCysArgGluMetLeuThrGly 40  
 |||||  
 Db 65 GAATTAAGGAAAAAGAAAGAAAGAAAGAGAGAGAGAGAAATCCAGGCCAATTGTGGC 124

QY 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisTrpIleGln 60  
 |||||  
 Db 125 ATAGATTTTATCATATCTGGATTTTGGATTTTGGATTTTCTCATCTGGATTCAG 184

QY 61 GluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
 |||||  
 Db 185 GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTACCTGCAGGGAAATGTTAACGGGA 244

QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
 |||||  
 Db 245 GGCTGCCTTCCCTGGGCAACAAGAGAGCCACCTGGGAGGAGAAAGTGCAGC 295

## RESULT 2

US-09-738-973-442  
 ; Sequence 442, Application US/09738973  
 ; Patent No. US20020110563A1  
 ; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Fling, Steven P.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Indirias, Carol Yoseph  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Elliot, Mark  
 ; APPLICANT: Mannion, Jane  
 ; APPLICANT: Kalos, Michael D.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: 210121.475C9  
 ; CURRENT APPLICATION NUMBER: US/09/738,973  
 ; CURRENT FILING DATE: 2000-12-14  
 ; NUMBER OF SEQ ID NOS: 587  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 442

; LENGTH: 337

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-738-973-442

## Alignment Scores:

Pred. No.: 2,77e-62 Length: 337  
 Score: 532.00 Matches: 97  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0

US-09-854-133-586 (1-97) x US-09-738-973-442 (1-337)

QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
 |||||  
 Db 5 GAGGTTGAAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTGGAGACTCTGTCTCAACA 64

QY 21 GluLeuArgLysLysGluArgLysLysArgGluValThrCysArgGluMetLeuThrGly 40  
 |||||  
 Db 65 GAATTAAGGAAAAAGAAAGAAAGAAAGAGAGAGAGAGAAATCCAGGCCAATTGTGGC 124

QY 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisTrpIleGln 60  
 |||||  
 Db 125 ATAGATTTTATCATATCTGGATTTTGGATTTTGGATTTTCTCATCTGGATTCAG 184

QY 61 GluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
 |||||  
 Db 185 GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTACCTGCAGGGAAATGTTAACGGGA 244

QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
 |||||  
 Db 245 GGCTGCCTTCCCTGGGCAACAAGAGAGCCACCTGGGAGGAGAAAGTGCAGC 295

## RESULT 3

US-09-854-133-441  
 ; Sequence 441, Application US/09854133  
 ; Publication No. US20020183499A1  
 ; GENERAL INFORMATION:

; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Secrist, Heather

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: 210121.475C10  
 ; CURRENT APPLICATION NUMBER: US/09/854,133

; CURRENT FILING DATE: 2001-05-11

; NUMBER OF SEQ ID NOS: 735

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 441

; LENGTH: 5981

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-854-133-441

## Alignment Scores:

Pred. No.: 6,32e-60 Length: 5981  
 Score: 527.00 Matches: 96  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.06% Indels: 0  
 DB: 9 Gaps: 0

US-09-854-133-586 (1-97) x US-09-854-133-441 (1-5981)

QY 2 ValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThrGlu 21  
 |||||  
 Db 3 GTTGAAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTGGAGACTCTGTCTCAACAGAA 62

QY 22 LeuArgLysLysGluArgLysLysArgGluValThrCysArgGluMetLeuThrGly 41  
 |||||  
 Db 63 TTAAGGAAAAAGAAAGAAAGAAAGAGAGAGAGAGAAATCCAGGCCAATTGTGGCATA 122

QY 42 AspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisTrpIleGlnGlu 61  
 |||||  
 Db 123 GATTTTATCATATCTGGATTTTGGATTTTGGATTTTCTCATCTGGATTCAGGAA 182

QY 62 SerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGly 81  
 |||||  
 Db 183 AGCCTGTTGTGTCCACCATCTCCAAAGGAGGTACTGCAGGGAAATGTTAACGGGAGGC 242

QY 82 CysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
 |||||

Db 243 TGCCTTCCCTGGGCAACAAGAGAGCCACCTGGGAGGAGAAAGTGCAGC 290

## RESULT 4

US-09-738-973-441  
 ; Sequence 441, Application US/09738973  
 ; Patent No. US20020110563A1  
 ; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Lodes, Michael J.



;; APPLICANT: Fling, Steven P.  
;; APPLICANT: Mohamath, Raodoh  
;; APPLICANT: Algate, Paul A.  
;; APPLICANT: Secrist, Heather  
;; APPLICANT: Indirias, Carol Yoseph  
;; APPLICANT: Benson, Darin R.  
;; APPLICANT: Elliot, Mark  
;; APPLICANT: Mannion, Jane  
;; APPLICANT: Kalos, Michael D.  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
;; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
;; FILE REFERENCE: 210121.475C9  
;; CURRENT APPLICATION NUMBER: US/09/738,973  
;; CURRENT FILING DATE: 2000-12-14  
;; NUMBER OF SEQ ID NOS: 587  
;; SOFTWARE: FastSEQ for Windows Version 3.0  
;; SEQ ID NO 441  
;; LENGTH: 5981  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-738-973-441

Alignment Scores:  
Pred. No.: 6.32e-60 Length: 5981  
Score: 527.00 Matches: 96  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.06% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-586 (1-97) x US-09-738-973-441 (1-5981)

QY 2 ValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThrGlu 21  
Db 3 GTTGAAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTGAGACTGTCTCAAACAGAA 62  
QY 22 LeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGlyIle 41  
Db 63 TTAAGGAAAAAGAAAGAAAGAAAGAGAGAGAGAGAGAAATCCAGGCCAATGTGGCATA 122  
QY 42 AspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGlnGlu 61  
Db 123 GATTTATCATATTTCTGGATTTTGGATTTCTTTTCTCATCACTGGATTCAGGAA 182  
QY 62 SerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGlyGly 81  
Db 183 AGCCTGTGTGTCCACCATCTCCAAAGAGAGTTACCTGCAGGGAAATGTTAACGGGAGGC 242  
QY 82 CysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
Db 243 TGCCTTCCCTGGGCAACAAGAGAGCCACCTGGGAGAGAGAAAGTGCAGC 290

RESULT 5  
US-09-854-133-440  
; Sequence 440, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 440  
; LENGTH: 2239  
; TYPE: DNA  
; ORGANISM: Homo sapiens

US-09-854-133-440  
Alignment Scores:  
Pred. No.: 1.24e-58 Length: 2239  
Score: 513.00 Matches: 97  
Percent Similarity: 84.35% Conservative: 0  
Best Local Similarity: 84.35% Mismatches: 0  
Query Match: 96.43% Indels: 18  
DB: 9 Gaps: 1

US-09-854-133-586 (1-97) x US-09-854-133-440 (1-2239)

QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
Db 2 GAGGTTGAAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTGAGACTGTCTCAAACA 61  
QY 21 GluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 40  
Db 62 GAATTAAGGAAAAAGAAAGAAAGAGAGAGAGAGAAATCCAGGCCAATGTGGC 121  
QY 41 IleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGln 60  
Db 122 ATAGATTTATCATATTTCTGGATTTTGGATTTCTTTTCTCATCACTGGATTCAG 181  
QY 61 GluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
Db 182 GAAAGCCTGTGTGTCCACCATCTCCAAAGAGAGTTACCTGCAGGGAAATGTTAACGGGA 241  
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArg----- 93  
Db 242 GGCTGCCTTCCCTGGGCAACAAGAGAGCCACCTGGGAGAGAGAGAGAGAGAGAGAG 301  
QY 94 -----ArgLysCysSer 97  
Db 302 CCTTTTCAGGAAGAGAGCGCCTTTTCAGGAAGAGAGAAAGTGCAGC 346

RESULT 6  
US-09-738-973-440  
; Sequence 440, Application US/09738973  
; Patent No. US20020110563A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Indirias, Carol Yoseph  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Elliot, Mark  
; APPLICANT: Mannion, Jane  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C9  
; CURRENT APPLICATION NUMBER: US/09/738,973  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 587  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 440  
; LENGTH: 2239  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-738-973-440

Alignment Scores:  
Pred. No.: 1.24e-58 Length: 2239  
Score: 513.00 Matches: 97  
Percent Similarity: 84.35% Conservative: 0  
Best Local Similarity: 84.35% Mismatches: 0  
Query Match: 96.43% Indels: 18  
DB: 10 Gaps: 1

US-09-854-133-586 (1-97) x US-09-738-973-440 (1-2239)

QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
Db 2 GAGGTGAAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTGAGACTCTGTCTCAACA 61

QY 21 GluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 40  
Db 62 GAATTAAGGAAAAAAGAAAGAAAGAGAGAGAGAGAAATCCAGGCCAATGTGGC 121

QY 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisTrpIleGln 60  
Db 122 ATAGATTTATCATATTCTGGATTTTGGATTCTTTGTTTCTCATCACTGGATTGAG 181

QY 61 GluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
Db 182 GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACCTGCAGGAAATGTTAACGGGA 241

QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArg----- 93  
Db 242 GCCTGCCTTCCCTGGGCAACAGAGCCACCTGGGCAGGACCGCTTTTCAGGAAGAGACG 301

QY 94 -----ArgLysCysSer 97  
Db 302 CCTTTTCAGGAAGAGACGCCCTTTTCAGGAAGAGAGAAAGTGCGAGC 346

RESULT 7

US-10-046-935-1307

; Sequence 1307, Application US/10046935

; Patent No. US20020156011A1

; GENERAL INFORMATION:

; APPLICANT: Jiang, Yuqiu

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Secrist, Heather

; APPLICANT: Wang, Aijun

; APPLICANT: Stolk, John A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER

; FILE REFERENCE: 210121.527C1

; CURRENT APPLICATION NUMBER: US/10/046,935

; CURRENT FILING DATE: 2002-01-15

; NUMBER OF SEQ ID NOS: 2239

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1307

; LENGTH: 572

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 9, 19, 461, 497, 500, 502

; OTHER INFORMATION: n = A,T,C or G

US-10-046-935-1307

Alignment Scores:

Pred. No.:	2.04e-20	Length:	572
Score:	221.00	Matches:	47
Percent Similarity:	70.00%	Conservative:	9
Best Local Similarity:	58.75%	Mismatches:	18
Query Match:	41.54%	Indels:	6
DB:	9	Gaps:	2

US-09-854-133-586 (1-97) x US-10-046-935-1307 (1-572)

QY 19 GlnThrGluLeuArgLysLysGluArgLysLysPheGlnAla 37  
Db 142 CAGTCTGAAGCAGAGGAGACATCGATCAGTAACACCAAGAGACACCAAGAGTGAAGT 201

QY 38 AsnCysGlyIleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHis 57  
Db 202 TTGTTTCTTCCCTCTGTTTATTTTCCCGTGTGTCCCTACTATGTGCA----- 255

QY 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77

Db 256 -----GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACCTGCAGGAAATG 306

QY 78 LeuThrGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97  
Db 307 TTAACGGGAGGCTGCCTTCCCTGGGCAACAAGAGAGCCACCTGGGCAGGAGAAAGTGCAGC 366

RESULT 8

US-09-878-178-1307

; Sequence 1307, Application US/09878178

; Patent No. US20020177552A1

; GENERAL INFORMATION:

; APPLICANT: Jiang, Yuqiu

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Secrist, Heather

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER

; FILE REFERENCE: 210121.527

; CURRENT APPLICATION NUMBER: US/09/878,178

; CURRENT FILING DATE: 2001-06-08

; NUMBER OF SEQ ID NOS: 2237

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1307

; LENGTH: 572

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(572)

; OTHER INFORMATION: n = A,T,C or G

US-09-878-178-1307

Alignment Scores:

Pred. No.:	2.04e-20	Length:	572
Score:	221.00	Matches:	47
Percent Similarity:	70.00%	Conservative:	9
Best Local Similarity:	58.75%	Mismatches:	18
Query Match:	41.54%	Indels:	6
DB:	9	Gaps:	2

US-09-854-133-586 (1-97) x US-09-878-178-1307 (1-572)

QY 19 GlnThrGluLeuArgLysLysGluArgLysLysPheGlnAla 37  
Db 142 CAGTCTGAAGCAGAGGAGACATCGATCAGTAACACCAAGAGACACCAAGTGAAGT 201

QY 38 AsnCysGlyIleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHis 57  
Db 202 TTGTTTCTTCCCTCTGTTTATTTTCCCGTGTGTCCCTACTATGTGCA----- 255

QY 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77  
Db 256 -----GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACCTGCAGGAAATG 306

QY 78 LeuThrGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97  
Db 307 TTAACGGGAGGCTGCCTTCCCTGGGCAACAAGAGAGCCACCTGGGCAGGAGAAAGTGCAGC 366

RESULT 9

US-10-146-502-1307

; Sequence 1307, Application US/10146502

; Publication No. US20030069180A1

; GENERAL INFORMATION:

; APPLICANT: Jiang, Yuqiu

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Secrist, Heather

; APPLICANT: Wang, Aijun

; APPLICANT: Stolk, John A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER

; FILE REFERENCE: 210121.527C2

; CURRENT APPLICATION NUMBER: US/10/146,502

; CURRENT FILING DATE: 2002-05-14



PRIOR APPLICATION NUMBER:

81 TTAACGG

81 TTAACGG



Tue May 13 12:12:53 2003

RESULT 15  
US-10-163-866-52  
; Sequence 52, Application US/10163866  
; Publication No. US20030027188A1  
; GENERAL INFORMATION:  
; APPLICANT: EXELIXIS, INC.  
; TITLE OF INVENTION: SLC7s AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE  
; FILE REFERENCE: EX02-080C  
; CURRENT APPLICATION NUMBER: US/10/163,866  
; CURRENT FILING DATE: 2002-06-05  
; PRIOR APPLICATION NUMBER: US 60/296,076  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/328,605  
; PRIOR FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: US 60/338,733  
; PRIOR FILING DATE: 2001-10-22  
; PRIOR APPLICATION NUMBER: US 60/357,253  
; PRIOR FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: US 60/357,600  
; PRIOR FILING DATE: 2002-02-15  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 52  
; LENGTH: 1528  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-163-866-52

Alignment Scores:  
Pred. No.: 1.95e-18 Length: 1528  
Score: 210.50 Matches: 38  
Percent Similarity: 95.00% Conservative: 0  
Best Local Similarity: 95.00% Mismatches: 1  
Query Match: 39.57% Indels: 1  
DB: 9 Gaps: 1

US-09-854-133-586 (1-97) x US-10-163-866-52 (1-1528)  
QY 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77  
Db 24 TGG---TCAGAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACCTGCAGGGAATG 80  
QY 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
Db 81 TTAACGGGAGGCTGCCTTCCTGGGCAACAAAGGAGGCCACCTGGGCAGGAGAAAGTGCAGC 140

Search completed: May 11, 2003, 16:28:53  
Job time : 120.027 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run On: May 11, 2003, 14:07:50 ; Search time 1470.45 Seconds  
(without alignments)  
1068.354 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHASLGDSSTLSQT.....LTGGCLPWATRSHLGRKCS\*97

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/SUFFIX=1st -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-DB=EST -QFMT=fastap -SUFFIX=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-UNITS=bits -START=1 -SUFFIX=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn\_1\_1\_2013@runat\_05052003\_173956\_349 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	221	41.5	936	12	BG284503	BG284503 602408645
2	221	41.5	1072	12	BG388107	BG388107 602413070
C 3	205	38.5	557	17	AQ002318	AQ002318 CIT-HSP-2
C 4	197.5	37.1	123	9	AI313891	AI313891 hpi-7 PMA
5	194	36.5	910	12	BG326527	BG326527 602425373
6	144.5	27.2	633	10	BB630988	BB630988 BB630988
7	144.5	27.2	652	10	BB612892	BB612892 BB612892
8	137.5	25.8	680	10	BB630675	BB630675 BB630675
9	131	24.6	524	12	BF286052	BF286052 EST450643
10	109.5	20.6	492	17	BH59880	BH59880 15_381443
11	109	20.5	669	17	AG037205	AG037205 Pan trogl
12	108.5	20.4	649	17	AQ629890	AQ629890 RPCI-11-4
C 13	107.5	20.2	599	17	AQ541929	AQ541929 RPCI-11-3
C 14	104.5	19.6	430	17	AQ414005	AQ414005 RPCI-11-1
15	102	19.2	919	14	BQ277344	BQ277344 AGENCOURT
C 16	101.5	19.1	492	17	AQ008942	AQ008942 RPCI11-21
C 17	100.5	18.9	777	17	AF102045	AF102045 AF102045
C 18	100.5	18.9	832	13	BM011880	BM011880 603636269
19	100	18.8	323	9	AA663028	AA663028 ab72b02.s
C 20	99.5	18.7	404	17	AQ596495	AQ596495 HS_5191_A
C 21	99.5	18.7	531	17	AQ822937	AQ822937 HS_3232_B
C 22	99	18.6	688	17	AG087340	AG087340 Pan trogl
C 23	98.5	18.5	708	17	AG119431	AG119431 Pan trogl
24	98	18.4	828	12	BG281385	BG281385 602401884
25	97.5	18.3	677	17	AG147658	AG147658 Pan trogl
C 26	96.5	18.1	562	17	AQ792364	AQ792364 HS_5255_B
C 27	96	18.0	298	12	BF886615	BF886615 RC6-TN007
C 28	96	18.0	673	9	AL696209	AL696209 DKFZp686C
C 29	95.5	18.0	518	10	AW849972	AW849972 IL3-CT021
30	95.5	18.0	586	10	AV764383	AV764383 AV764383
31	95	17.9	350	12	BF095483	BF095483 IL2-UT007
32	95	17.9	501	17	B99029	B99029 CIT-HSP-228
33	93.5	17.6	226	10	AW798344	AW798344 RC1-UM004
C 34	93.5	17.6	917	12	BE733566	BE733566 601565919
C 35	93	17.5	290	12	BF927188	BF927188 CM4-NT024
36	93	17.5	599	17	AQ626336	AQ626336 CITBI-E1-
37	93	17.5	649	17	AG038593	AG038593 Pan trogl
C 38	93	17.5	1112	13	BM450518	BM450518 AGENCOURT
39	92.5	17.4	258	12	BF959970	BF959970 QV2-NN004
40	92.5	17.4	716	12	BF676985	BF676985 602084212
41	92	17.3	417	9	AA424117	AA424117 zv80el2.r
42	91.5	17.2	502	9	AA458705	AA458705 aa87g08.s
C 43	91	17.1	489	9	AA780515	AA780515 ac71d09.s
44	91	17.1	1006	12	BG707662	BG707662 602670550
45	90.5	17.0	385	17	AQ029438	AQ029438 RPCI11-39

ALIGNMENTS

RESULT 1  
BG284503  
LOCUS BG284503 936 bp mRNA linear EST 21-FEB-2001  
DEFINITION 602408645F1 NIH\_MGC\_91 Homo sapiens CDNA clone IMAGE:4537810 5',  
mRNA sequence.  
ACCESSION BG284503  
VERSION BG284503.1 GI:13035516  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 936)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)







Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsic.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura  
, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y. Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

**FEATURES**  
**source**

```

e mouse tissues.
Location/Qualifiers
1. .633
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A130046A05"
/clone_lib="RIKEN full-length enriched, 16 days neonate
thymus"
/tissue_type="thymus"
/dev_stage="16 days neonate"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAGAGCTCTTTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCGAGTTAATTAATCCCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FIC I."

```

144 a 154 c 175 g 159 t 1 others

RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAAGGATCCAGAGCTCTTTT TTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adaptor of sequence [5' GAGAGAGAGATTCTCGAGTAATAAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

	154 c	175 g	159 t	1 others
144 a				

cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTAAATTAATATCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I." 144 a 154 c 175 g 159 t 1 others

strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAATTAATTCCTCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda

sequence [5' GAGAGAGAGATTCTCGAGTTAATTAATTAATCCCCCCCCCCC  
3']. cDNA was cleaved with *Xho*I and *Bam*HI. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda  
FLC I.<sup>a</sup>

144 a 154 c 175 g 159 t 1 others

	144 a	154 c	175 g	159 t	1 others
CS: CS-99 was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."					

FLC I."	154 c	175 g	159 t	1 others
144 a				

FLC I."	175 g	159 t	1 others
144 a	154 c		

144 a	154 c	175 g	159 t	1 others
-------	-------	-------	-------	----------









Alignment Scores:		
Pred. No.:	0.926	Length: 492
Score:	109.50	Matches: 37
Percent Similarity:	47.31%	Conservative: 7
Best Local Similarity:	39.78%	Mismatches: 36
Query Match:	20.58%	Indels: 14
DB:	17	Gaps: 3

BASE COUNT	185 a	161 c	173 g	150 t
ORIGIN				

Alignment Scores:	
Pred. No.:	1.31
Score:	109.00
Percent Similarity:	60.27%
Best Local Similarity:	47.95%
Mismatches:	11
Conservative:	9
Matches:	35
Length:	669





Search completed: May 11, 2003, 15:28:01  
Job time : 1476.45 secs



29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rnd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	532	100.0	337	6	AX321911	AX321911 Sequence
2	523	98.3	2239	6	AX321909	AX321909 Sequence
3	339.5	63.8	1683	9	AB042201	AB042201 Homo sapi
4	339.5	63.8	1861	9	AB026891	AB026891 Homo sapi
5	339.5	63.8	1874	9	AF200708	AF200708 Homo sapi
6	339.5	63.8	2482	9	AF252872	AF252872 Homo sapi
7	339	63.7	575	6	AX341060	AX341060 Sequence
8	327.5	61.6	596	6	AX351341	AX351341 Sequence
9	307.5	57.8	2155	9	BC012087	BC012087 Homo sapi
10	297	55.8	927	10	AY065541	AY065541 Mus muscu
11	297	55.8	4845	9	HUMEDHB17	M27138 Human estro
12	296.5	55.7	2000	9	AB040875	AB040875 Homo sapi
13	294	55.3	1666	9	AC104107	AC104107 Homo sapi
14	293.5	55.2	1671	10	BC022180	BC022180 Mus muscu
15	293	55.1	4068	10	AF230072	AF230072 Rattus no
16	292.5	55.0	3995	6	AX375015	AX375015 Sequence
17	292.5	55.0	3995	9	HSU83192	U83192 Homo sapien
18	291	54.7	4623	10	BC034547	BC034547 Mus muscu
19	290	54.5	1600	8	AF019376	AF019376 Brassica
20	289.5	54.4	3994	9	AL750302	AL750302 Human DNA
21	289	54.3	2382	9	AK023017	AK023017 Homo sapi
22	288	54.1	1269	3	AF160879	AF160879 Drosophil
23	288	54.1	2385	9	HS300588	AJ300588 Homo sapi
24	287.5	54.0	2427	6	AX001017	AX001017 Sequence
25	287.5	54.0	2427	6	BD005487	BD005487 G-coupled
26	287	53.9	900	9	HSGCKRIN1	Y09593 H.sapiens G
27	286.5	53.9	4314	9	HSM804205	AL832894 Homo sapi
28	286	53.8	4363	9	HSM804473	AL833162 Homo sapi
29	285.5	53.7	1906	4	AF339909	AF339909 Capra hir
30	285.5	53.7	3848	9	HSL81672	L81672 Homo sapien
31	285	53.6	3608	10	BC028880	BC028880 Mus muscu
32	284.5	53.5	938	10	MMU04333	U04333 Mus musculu
33	284.5	53.5	1114	9	HUMF13A01	M21986 Human coagu
34	284.5	53.5	1811	9	AK025185	AK025185 Homo sapi
35	284.5	53.5	2008	9	AK023483	AK023483 Homo sapi
36	284.5	53.5	2727	9	AK001575	AK001575 Homo sapi
37	284.5	53.5	4313	9	HSM800932	AL117425 Homo sapi
38	284	53.4	1310	17	AF116717	AF116717 Homo sapi
39	284	53.4	2723	9	AK027045	AK027045 Homo sapi
40	284	53.4	3267	9	BC012925	BC012925 Homo sapi
41	284	53.4	3890	14	RNMRS1	J04357 Red clover
42	283.5	53.3	1517	8	AY038318	AY038318 Inocybe l
43	283	53.2	2533	6	AR122536	AR122536 Sequence
44	283	53.2	2590	9	AB061850S1	AB061850 Homo sapi
45	283	53.2	2599	10	AF068707	AF068707 Mus muscu

ALIGNMENTS

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 18:33:52 ; Search time 1548 Seconds  
(without alignments)  
1823.625 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHASLGDSSETLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62  
Xgapop 0.0 , Xgapext 0.5  
Ygapop 0.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0-

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 3813070

Minimum DB seq length: 0  
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -MODEL=frame+ -p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_174426\_1210/app\_query.fasta\_1.263  
-DB=GenEmbl -Qfmt=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=5000  
-USER=US09854133@cgn\_1\_1\_1687\_@runat\_05052003\_174426\_1210 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=0 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=0 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pi:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_in:\*  
18: em\_mu:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

```

AX321911
LOCUS      AX321911
DEFINITION Sequence 442 from Patent WO0172295.
ACCESSION  AX321911
VERSION     AX321911.1  GI:17906521
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,
            Indrias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,
            Mannion,J. and Kalos,M.D.
TITLE       Compositions and methods for the therapy and diagnosis of lung
            cancer
JOURNAL
Patent: WO 0172295-A 442 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES   source
            1. .337
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT 103 a 60 c 93 g 81 t
ORIGIN
Alignment Scores:
Pred. No.: 2.97e-06
Score: 532.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 6
Length: 337
Matches: 97
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0
US-09-854-133-586 (1-97) x AX321911 (1-337)
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20
Db 5 GAGGTTGAAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTGAGACTCTGTCTCAACA 64
QY 21 GluLeuArgLysLysGluArgLysLysLysLysLysLysLysLysLysLysLysLysLys 40
Db 65 GAATTAAGGAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 40
QY 41 IleAspPheIleIlePheTrpIlePheTrpIlePheTrpIlePheTrpIlePheTrpIle 60
Db 125 ATAGATTTTATCATATTTCTGGATTTTCTGGATTTTCTGGATTTTCTGGATTTTCTGG 184
QY 61 GluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGly 80
Db 185 GAAAGCCTGTGTGTCCACCATCTCCAAAGAGAGGTTACCTGCAGGAAATGTTAAGCGGA 244
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97
Db 245 GGCTGCCTTCCCTGGGCAACAAGGAGCCACCTGGGCAAGGAAAGTGCAGC 295
RESULT 2
AX321909
LOCUS      AX321909
DEFINITION Sequence 440 from Patent WO0172295.
ACCESSION  AX321909
VERSION     AX321909.1  GI:17906515
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,
            Indrias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,
            Mannion,J. and Kalos,M.D.
TITLE       Compositions and methods for the therapy and diagnosis of lung
            cancer
JOURNAL
Patent: WO 0172295-A 440 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES   source
            1. .337
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT 103 a 60 c 93 g 81 t
ORIGIN
Alignment Scores:
Pred. No.: 2.97e-06
Score: 532.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 6
Length: 337
Matches: 97
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0
US-09-854-133-586 (1-97) x AX321911 (1-337)
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20
Db 5 GAGGTTGAAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTGAGACTCTGTCTCAACA 64
QY 21 GluLeuArgLysLysGluArgLysLysLysLysLysLysLysLysLysLysLysLysLys 40
Db 65 GAATTAAGGAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 40
QY 41 IleAspPheIleIlePheTrpIlePheTrpIlePheTrpIlePheTrpIlePheTrpIle 60
Db 125 ATAGATTTTATCATATTTCTGGATTTTCTGGATTTTCTGGATTTTCTGGATTTTCTGG 184
QY 61 GluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGly 80
Db 185 GAAAGCCTGTGTGTCCACCATCTCCAAAGAGAGGTTACCTGCAGGAAATGTTAAGCGGA 244
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97
Db 245 GGCTGCCTTCCCTGGGCAACAAGGAGCCACCTGGGCAAGGAAAGTGCAGC 295
RESULT 2
AX321909
LOCUS      AX321909
DEFINITION Sequence 440 from Patent WO0172295.
ACCESSION  AX321909
VERSION     AX321909.1  GI:17906515
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,
            Indrias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,
            Mannion,J. and Kalos,M.D.
TITLE       Compositions and methods for the therapy and diagnosis of lung
            cancer
JOURNAL
Patent: WO 0172295-A 440 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES   source
            1. .337
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT 103 a 60 c 93 g 81 t
ORIGIN
Alignment Scores:
Pred. No.: 2.97e-06
Score: 532.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 6
Length: 337
Matches: 97
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0
US-09-854-133-586 (1-97) x AX321911 (1-337)
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20
Db 5 GAGGTTGAAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTGAGACTCTGTCTCAACA 64
QY 21 GluLeuArgLysLysGluArgLysLysLysLysLysLysLysLysLysLysLysLysLys 40
Db 65 GAATTAAGGAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 40
QY 41 IleAspPheIleIlePheTrpIlePheTrpIlePheTrpIlePheTrpIlePheTrpIle 60
Db 125 ATAGATTTTATCATATTTCTGGATTTTCTGGATTTTCTGGATTTTCTGGATTTTCTGG 184
QY 61 GluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGly 80
Db 185 GAAAGCCTGTGTGTCCACCATCTCCAAAGAGAGGTTACCTGCAGGAAATGTTAAGCGGA 244
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97
Db 245 GGCTGCCTTCCCTGGGCAACAAGGAGCCACCTGGGCAAGGAAAGTGCAGC 295

```

```

CORIXA CORPORATION (US)
FEATURES   source
            1. .2239
            Location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT 619 a 444 c 493 g 683 t
ORIGIN
Alignment Scores:
Pred. No.: 0.000752
Score: 523.00
Percent Similarity: 84.35%
Best Local Similarity: 84.35%
Query Match: 98.31%
DB: 6
Length: 2239
Matches: 97
Conservative: 0
Mismatch: 0
Indels: 18
Gaps: 1
US-09-854-133-586 (1-97) x AX321909 (1-2239)
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20
Db 2 GAGGTTGAAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTGAGACTCTGTCTCAACA 61
QY 21 GluLeuArgLysLysGluArgLysLysLysLysLysLysLysLysLysLysLysLysLys 40
Db 62 GAATTAAGGAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 40
QY 41 IleAspPheIleIlePheTrpIlePheTrpIlePheTrpIlePheTrpIlePheTrpIle 60
Db 122 ATAGATTTTATCATATTTCTGGATTTTCTGGATTTTCTGGATTTTCTGGATTTTCTGG 181
QY 61 GluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGly 80
Db 182 GAAAGCCTGTGTGTCCACCATCTCCAAAGAGAGGTTACCTGCAGGAAATGTTAAGCGGA 241
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArg 93
Db 242 GGCTGCCTTCCCTGGGCAACAAGGAGCCACCTGGGCAAGGAAAGTGCAGC 301
QY 94 -----ArgLysCysSer 97
Db 302 CCTTTTCAGGAAGAGAGAGCGCTTTTCAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 346
RESULT 3
AB042201
LOCUS      AB042201
DEFINITION Homo sapiens xCT gene for cystine/glutamate transporter, exon 1, partial cds.
ACCESSION  AB042201
VERSION     AB042201.1  GI:7670267
KEYWORDS    cystine/glutamate transporter.
SOURCE      Homo sapiens DNA.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
AUTHORS     Sato,H., Tamba,M., Kuriyama-Matsumura,K., Okuno,S. and Bannai,S.
TITLE       Molecular cloning and expression of human xCT, the light chain of
            amino acid transport system xc-
JOURNAL     Antioxidants and Redox Signaling 2, 665-671 (2000)
REFERENCE   2 (bases 1 to 1683)
AUTHORS     Sato,H. and Bannai,S.
TITLE       Direct Submission
JOURNAL     Submitted (24-APR-2000) Hideyo Sato, University of Tsukuba,
            Institute of Basic Medical Sciences; Tennodai 1-1-1, Tsukuba,
            Ibaraki 305-8575, Japan (E-mail:hideyo-smd.tsukuba.ac.jp,
            Tel:81-298-53-3282, Fax:81-298-53-3039)
FEATURES   Location/Qualifiers
            1. .1683
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1001. .1683
            /gene="xCT"
            1001. .1555
            /exon

```





ACCESSION AF200708  
VERSION AF200708.1 GI:11493651  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens.

REFERENCE 1 (bases 1 to 1874)  
AUTHORS Conklin,D.S. and Beach,D.H.  
TITLE CCBRL1, novel CD98 light chain implicated in redox control and calcium signaling

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1874)  
AUTHORS Conklin,D.S. and Beach,D.H.  
TITLE Direct Submission  
JOURNAL Submitted (01-NOV-1999) Cold Spring Harbor Labs, 1 Bungtown Rd, Cold Spring Harbor, NY 11724, USA

FEATURES  
source Location/Qualifiers  
1..1874

CDS  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="fibrosarcoma HT1080"  
246..1751

/note="x-c transporter; CD98 light chain"  
/codon\_start=1  
/product="calcium channel blocker resistance protein CCBRL1"  
/protein\_id="AAG35592.1"  
/db\_xref="GI:11493652"

/translation="MVRKPVVSTISKGYLQGNVNGRLPSLGNKEPPQEKVQLKRV  
TLLRGVSIIGTIIGAGIFISPKGVLTQNTSGVSLTIWTCVGLSFLGALSVAELGT  
PELAIKLITAVGITVVMVLSMSVSWARIQIFLTFCCKLTAILIIIVPGVMQLIKGT  
QNFKDAFSGRDSITRLPLAFYGYMYAGWFLNFVTEVENPEKTIPLAICISMAI  
VTIGVLTNVAFTTINAEELLSNAVAVTFSERLNGFSLAVPIFVALSCFGSMNGG  
VFAVSRFLYVASREGLHPEILSMIHKHTPLPAVIVLHPLTMIMLFSGLDLSLLNFL  
SFARWLFGLAVAGLIYLRKCPDMHRPFKPLFIPALFSTCLFVVALSLYSDPFST  
GIGFVITLTGVPAYLYFIWDDKPRFRIMSEKITRTLTQIILEVVPEDKL"

BASE COUNT 490 a 407 c 422 g 555 t  
ORIGIN

## Alignment Scores:

Pred. No.: 446 Length: 1874  
Score: 339.50 Matches: 62  
Percent Similarity: 57.25% Conservative: 13  
Best Local Similarity: 47.33% Mismatches: 7  
Query Match: 63.82% Indels: 49  
DB: 9 Gaps: 28

US-09-854-133-586 (1-97) x AF200708 (1-1874)

QY 1 GluValGluValSerArg----AspHis---Ala-----Ser-----Leu 11  
Db 16 GAG---GAGGTGGAGAAATGAGAGCAGCATGATACACAGGTGTTCTGAGTAGTAATTA 72  
QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22  
Db 73 ---GATCGCTGTGAAGGAAACCAACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 126  
QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33  
Db 127 CGCTGAGAGACACAGTCTGAAAGCAGAGGAGATCGATCGATGATACACCAAGAGACACC 186  
QY 34 LysPheGlnAlaAsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---Phe 49  
Db 187 AAA-----GTTGAAAGTTT---GTTTCTTCCCTCTGTTTATTT 225  
QY 50 Trp-----Ile---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysPro 66  
Db 226 TTCCCCCGTGTGCTCCCTACTA-----TGG---TCAGAAAGCCTGTGTGTCCA 270  
QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla 86  
|||||

Db 271 CCATCTCCAAAGGAGGTTACCTGCAGGAAATGTTAACGGGAGGCTGCCTTCCCTGGGCA 330  
QY 87 ThrArgSerHisLeuGlyArgArgLysCysSer 97  
|||||

Db 331 ACAAGGAGCCACCTGGGCGAGGAGAAAGTGCAGC 363

RESULT 6

AF252872

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

Location/Qualifiers

1..2482

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/tissue\_type="placenta"

232..1737

/codon\_start=1

/product="cystine/glutamate transporter xCT"

/protein\_id="AAK49111.1"

/db\_xref="GI:13924720"

/translation="MVRKPVVSTISKGYLQGNVNGRLPSLGNKEPPQEKVQLKRV

TLLRGVSIIGTIIGAGIFISPKGVLTQNTSGVSLTIWTCVGLSFLGALSVAELGT

TIKSGGHYTYILEVFGPLPAFVRVWVLLIIRPAATAVISLAFGRYILEPFFIQCEI

PELAIKLITAVGITVVMVLSMSVSWARIQIFLTFCCKLTAILIIIVPGVMQLIKGT

QNFKDAFSGRDSITRLPLAFYGYMYAGWFLNFVTEVENPEKTIPLAICISMAI

VTIGVLTNVAFTTINAEELLSNAVAVTFSERLNGFSLAVPIFVALSCFGSMNGG

VFAVSRFLYVASREGLHPEILSMIHKHTPLPAVIVLHPLTMIMLFSGLDLSLLNFL

SFARWLFGLAVAGLIYLRKCPDMHRPFKPLFIPALFSTCLFVVALSLYSDPFST

GIGFVITLTGVPAYLYFIWDDKPRFRIMSEKITRTLTQIILEVVPEDKL"

BASE COUNT 735 a 484 c 521 g 742 t  
ORIGIN

Alignment Scores:

Pred. No.: 917 Length: 2482

Score: 339.50 Matches: 62

Percent Similarity: 57.25% Conservative: 13

Best Local Similarity: 47.33% Mismatches: 7

Query Match: 63.82% Indels: 49

DB: 9 Gaps: 28

US-09-854-133-586 (1-97) x AF252872 (1-2482)

QY 1 GluValGluValSerArg---AspHis---Ala-----Ser-----Leu 11

Db 2 GAG---GAGGTGGAGAAATGAGAGCAGCATGATACACAGGTGTTCTGAGTAGTAATTA 58

QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22

Db 59 ---GATCGCTGTGAAGGAAACCAACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 112

QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33

Db 113 CGCTGAGAGACACAGTCTGAAAGCAGAGGAGACATGATGATGATGATGATGATGATGAT 172

QY 34 LysPheGlnAlaAsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---Phe 49





US-09-854-133-586 (1-97) x BC012087 (1-2155)

CDs

```
/translation="MVLFQIKETFFEFMTGLNISVSDYFFRLLLEGVKKKPIHL
ILIHAFNTLILTKGFRKTVAAFGRLNFDLDIGCKSIYLERVARGLSICTSSLLT
```



```

intron      /number=2
2229. .2375
/note="EDHB17 intron B"
2376. .2555
/note="EDHB17 intron C"
2556. .2730
/note="EDHB17 intron D"
2731. .2824
/note="EDHB17 intron E"
2825. .3340
/note="EDHB17 intron F"
3341. .3518
/note="EDHB17 intron G"
3519. .3604
/note="EDHB17 intron H"
3605. .3874
/note="estradiol 17 beta-dehydrogenase"
/note="Alu repeat copy D"
4008. .4257
/note="Alu repeat copy E"
4516. .4780
/note="Alu repeat copy F"
1032 a 1341 c 1449 g 1023 t
ORIGIN      Chromosome 17q11-q12.

Alignment Scores:
Pred. No.: 1.21e+05
Score: 297.00
Percent Similarity: 33.75%
Best Local Similarity: 27.08%
Query Match: 55.83%
DB: 9

US-09-854-133-586 (1-97) x HUMEDHB17 (1-4845)
QY 1 Glu---ValGluValSer-----Arg-----Asp-----His 8
Db 4606 GAATCACTGAA---TCTGGGAAGCGGAGATTGTGGTGAGCCGAGATTGCCCATTCAC 4550
QY 9 AlaSerLeuGlyAsp---SerGluThrLeuSerGlnThrGluLeuArgLysLysGluArg 27
Db 4549 TCCAGCCTAGGCAACAGAGAGCGAACTGTCTCTCAG---AAATAAAAAAGAAAGA 4493
QY 28 LysLys---LysArg---Glu-----Arg----- 33
Db 4492 AAGAAAGAAAGAAAGAACTGAGAGAAAGCAAGAACTTGCCTGAGCTCATGGCTAATG 4433
QY 34 -----Lys-----PheGlnAlaAsn----- 38
Db 4432 GGGTCCAAGTCACAACTGACTCTACCCCAAGCCAAATCCACCACTCCTCCTGCAGA 4373
QY 39 Cys-----GlyIleAsp-----Phe-----Ile----- 44
Db 4372 TGCCCTTGGGTATCCAGAGACTGTGTCTGGCCTTTAATATGAGCTTGAGGTGATCCT 4313
QY 45 ---Ile-----PheTrp-----Ile-----Phe---Trp---Ile---Leu-Le 53
Db 4312 GGGCTGGCTTCTTCTGGACCTTGGTTTCTGAGTTTCTGGTTGGTTGGTTTCATTGATT 4253
QY 53 uphe-----Ser-----His---HisTrp-----IleGlnGluSerLeuLeuCy 65
Db 4252 AATTGACAGGGTCTCACTCTGTCAACCCAGGCTGGAGTGCAGTGGCCCCCATCA-----TG 4199
QY 65 s-----ProProSerPro-----LysGluValThrCysArg----- 75
Db 4198 CTCACCTGACCCCTCAACTCCTGGGCTCAAGAA---TGGTGGCCTGTGCTGATTTTTA 4142
QY 76 -----GluMet-----Leu-----ThrGly----- 80
Db 4141 AATTCTTTTATTATTGAGAGATGGGGTCTCACTGTGTGTCTGCTGGCTGCTCCAAC 4082
QY 81 ----Gly-----Cys-----LeuPro-----Trp----- 85
```

```

Db 4081 TCCGGGGCTTGAGCGATCCTCCTGCCTCCGCTCCCAAGCGCTGGCATACAGCGACA 4022
QY 86 -----Ala-ThrArg-----SerHis- 90
Db 4021 GCCATCGCGCTAGCTGCGCTGGTAAACTGGCTAACGCGATCTATCTTAATTAGCCACC 3962
QY 91 -----Leu-----Gly-----Arg-----Lys 95
Db 3961 CACAGCTGTACCGCGCGCCCATCCCGAGGACACACAGACCCAGGGGACAAAGA 3906
RESULT 12
AB040875 2000 bp mRNA linear PRI 03-APR-2001
LOCUS Homo sapiens hxCt mRNA for cysteine/glutamate exchanger, complete cds.
DEFINITION AB040875
ACCESSION AB040875
VERSION AB040875.1 GI:13516845
KEYWORDS
SOURCE
ORGANISM Homo sapiens adult cDNA to mRNA, clone_lib:brain cDNA library
clone:hxCt.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kim,J.Y., Chairoungdua,A., Cha,S.H., Segawa,H., Matsuo,H.,
TITLE Kim,D.K., Endou,H. and Kanai,Y.
JOURNAL Human cystine/glutamate exchanger: cDNA cloning and upregulation by
REFERENCE oxidative stress in glioma cells
AUTHORS Unpublished
TITLE 2 (bases 1 to 2000)
JOURNAL Kanai,Y.
Direct Submission
Submitted (01-APR-2000) Yoshikatsu Kanai, Kyorin University School
of Medicine, Department of Pharmacology and Toxicology; 6-20-2
Shinkawa, Mitaka, Tokyo 181-8611, Japan
(E-mail:ykanai@kyorin-u.ac.jp, Tel:+81-422-47-5511(ex.3453),
Fax:+81-422-79-1321)
FEATURES
Location/Qualifiers
1..2000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hxCt"
/clone_lib="brain cDNA library"
/dev_stage="adult"
1..2000
/gene="hxCt"
136..1620
/gene="hxCt"
/note="Na+-independent amino acid transporter
transmembrane protein"
/codon_start=1
/product="cystine/glutamate exchanger"
/protein_id="BAB40574.1"
/db_xref="GI:13516846"
/translation="MVRKPVVSTISKGYLQNVNGLPSLGNKEPPGQEKVQLKRV
TLRGVSIIGTIIGAGFISPRGVLQNTGSGMSLTITWVCGVLSLFGALSAYELGT
TIKSGGHVYIILEVFGPLPAFVRVWVLLIIRPAATAVISLAFGRYILEPFIQCEI
PELAIKLITAVGITVVMVLSMSVSWARIQIFLTFCCLTAIILIIIVPGVQLIKGQT
QNFDAFSGRDSITRLPLAFYGMAYAGWFLNFVTEVENPEKTIPLAICISMAI
VTIGVLTNVAYFTTINAEELISNAVAVTFSERLGNFSLAVPIFVALSCFGSMNGG
VFVSRFLFYVASREGHLPEILSMIHVKHTPLPAVIVLHPLTMIPLFSGDLSLLNFL
SFARWLFIGLAVAGLIYLRKCPDMHRPKVPLFIPALFSTCTCFMVALSLYSDPFST
GIGFVITLTGVPAYLYFIWDKKPRWFRIMSGFLALMPAQACDM"
BASE COUNT 505 a 418 c 435 g 642 t
ORIGIN
Alignment Scores:
Pred. No.: 1.32e+04
Score: 296.50
Percent Similarity: 58.88%
Best Local Similarity: 53.27%
Query Match: 55.73%
DB: 9
Length: 2000
Matches: 57
Conservative: 6
Mismatch: 10
Indels: 34
Gaps: 19
```



US-09-854-133-586 (1-97) x AB040875 (1-2000)

QY 2 ValGlu---ValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeu---SerGln 19  
Db 4 GTGAACACATATA-----GCG-----CTG-----AGAGAGACAGTCTGAAAGCAG 42

QY 20 ThrGluLeuArgLys---Lys---GluArgLysLysLysArg---GluArg---LysPhe 35  
Db 43 -----AGGAAGACATCGATCAGT---AACACCAAGAGACACCAAGAGTTGAAAGTTT 90

QY 36 GlnAlaAsnCysGlyIleAspPheIle---Ile---PheTrpIlePhe-----TrpIle 51  
Db 91 -----TGT-----TTTCTTTCCCTCTGTTT---ATTTTCCCTCCGTTGTC 129

QY 52 -LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysProProSerProLysGln 71  
Db 130 CTTACTA-----TGG---TCAGAAAGCCTGTTGTGTCACCATCTCCAAAGGA 174

QY 71 uValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLe 91  
Db 175 GGTACTCTGCAGGGAAATGTTAAAGGGAGGCTGCCTCCCTGGGCAACAAGGAGCCACCT 234

QY 91 uGlyArgArgLysCysSer 97  
Db 235 GGGCAGGAGAAAGTGCAGC 253

RESULT 13  
AC104107  
LOCUS Homo sapiens chromosome 19 clone CTC-473C7, complete sequence.  
DEFINITION AC104107  
ACCESSION AC104107.3 GI:21844562  
VERSION HTG.  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1666)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1666)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (04-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 3 (bases 1 to 1666)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (11-APR-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 4 (bases 1 to 1666)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (16-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Jul 16, 2002 this sequence version replaced gi:20136898.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www-shgc.stanford.edu  
Quality: Phrap Quality >=40 99.1% of Sequence;  
Estimated Total Number of Errors is 0.  
NOTE: This sequence is not the entire sequence of the clone. It is sequence generated to span the gap between AC008540 and AC008586. The overlap with AC008540 is 652bp and the overlap with AC008586 is 1018bp. The sequence was finished by the Stanford Human Genome Center and Los Alamos National Laboratory.  
Location/Qualifiers  
1. .1666  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 443 a 364 c 433 g 426 t  
ORIGIN  
Alignment Scores: 9.96e+03 Length: 1666  
Pred. No.: 294.00 Matches: 62  
Score: 31.73% Conservatives: 17  
Percent Similarity: 24.90% Mismatches: 12  
Best Local Similarity: 55.26% Indels: 159  
Query Match: 9 Gaps: 50  
DB:  
US-09-854-133-586 (1-97) x AC104107 (1-1666)  
QY 1 GluValGlu-----ValSerArgAspHis---AlaSerLeu-----Gly---AspSer 14  
Db 91 GAGCGGAGGTTGCAGTGAGCCCAAGATCATGCCGCTGCACCTAGCCTGGGTGACAGAGC 150  
QY 15 GluThrLeuSerGln---ThrGluLeuArgLysLysGlu---ArgLysLysLysArgGlu 32  
Db 151 GAGACCTTGCTCAAAAAA-----AAAAAGAAAGAAATATAAAAAAAGAA 207  
QY 33 ArgLys---PheGlnAla---Asn-----Cys-----Gly----- 40  
Db 208 ---AAAGAAATGGAATCATACATATCATACAGTCTGTCTCTTTTGTGATGGCTCCTTTC 264  
QY 41 -----Ile---Asp-----PheIleIle-----Phe--- 46  
Db 265 ACTCCACACTTGTGAGATGCATCCTCTCTG-ATCTTTTCATCAGAATCTGTCTCTAT 323  
QY 47 -----Trp-----Ile-----Phe-----Trp----- 50  
Db 324 TATACTCCAGTGTATGGGTGGACCACTGTTTGTCTCTCTTGTGAATGGGCATATGGT 383  
QY 51 Ile---LeuLeuPhe-----Ser-----His----- 56  
Db 384 TTGTTTCTAGTATTGGCTAACATGAATAAAGCTGCTATGAACATCCATGTATAAAGTC 443  
QY 57 -----His-----His----- 57  
Db 444 TTTGTGGGGCCAGGCTCAGTGGCTCAGCCCTATATATCCAGCACTTTGGGAGGCCAAG 503  
QY 58 Trp-----IleGlnGluSerLeuLeuCysPro----- 66  
Db 504 TGGGAAGATCACCTGAGGTGAGGTCAAGACCAG-----CCAGGCCAACATGGT 554  
QY 67 -----ProSer---ProLys---Glu-----Val-----ThrCys----- 74  
Db 555 GAAACCCCGTCTCTACAAAAAATACAAAAATAGCCAGCCAGGATGGTGCATGCCCTGTAA 614  
QY 75 -----Arg---Glu-----Met---LeuThrGly-----Gly 81  
Db 615 TCTCAGCTACTCAGGAGGCTGAGGTGGGAGAAATCAGTTGAACCTG---GGAGGTGGAGT 671  
QY 82 Cys-----Leu-----Pro---TrpAlaThr---Arg----- 88  
Db 672 TGCAGTGTGCTGAAATTTGCACCACCTGCACCTCCAGCTGGGCAACAGAGCGAGGCTCTGTC 731  
QY 89 Ser---His-----Leu---Gly-----Arg---Arg 94  
Db 732 TCAAAAAACAAACAAACAAATGCTTTTGTGGGATCTCTGGGTAATACAGAGAGCGG 791  
QY 95 LysCys-----Ser 97  
Db 792 AATTGTTGAGTCACCTGGTGAGTGTCC 818  
RESULT 14  
BC022180/c BC022180 1671 bp mRNA linear ROD 07-AUG-2002  
LOCUS Mus musculus, Similar to  
DEFINITION UDP-N-acetyl-alpha-D-galactosamine:(N-acetylneuraminyl)-  
galactosylglucosylceramide-beta-1,

## AUTHORS

Ohto, T., Uchida, H.

1 (bases 1 to 4068)  
Ohto, T., Uchida, H., Yamazaki, H., Keino-Masu, K., Matsui, A. and Masu, M.

100





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run On: May 11, 2003, 17:20:55 ; Search time 156 Seconds  
(without alignments)  
1400.281 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHASLGDSLTSLT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62  
Xgapop 0.0 , Xgapext 0.5  
Ygapop 0.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4313688

Minimum DB seq length: 0  
Maximum DB seq length: 5000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_174425\_1202/app\_query.fasta\_1.263  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=5000  
-USER=US09854133@cgn\_1\_1\_200\_@runat\_05052003\_174425\_1202 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=0 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=0 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	532	100.0	337	23	AAD23462	Human lung tumour-
2	523	98.3	2239	23	AAD23460	Human lung tumour-
3	339.5	63.8	800	20	AAZ16609	Human gene express
4	339	63.7	575	24	ABL37718	Human colon tumour
5	327.5	61.6	596	24	ABK27651	Human colon cancer
6	304	57.1	2752	22	ABA16566	Human nervous syst
7	298.5	56.1	1687	22	ABA07916	Human ovarian and
8	298.5	56.1	1687	22	AAO3733	Human reproductive
9	298.5	56.1	1688	22	ABA07917	Human ovarian and
10	298.5	56.1	1688	22	AAO3734	Human reproductive
11	292.5	55.0	3995	24	ABK35489	Human endometrial
12	292.5	55.0	3995	24	ABK35548	Gene DLG4 differen
13	292	54.9	2197	22	ABA07287	Human pancreatic c
14	292	54.9	2197	22	AAK89932	Human digestive sy
15	291	54.7	2751	22	ABA16568	Human nervous syst
16	289.5	54.4	2592	23	ABL14427	Drosophila melanog
17	289	54.3	2382	22	AAH17785	Human cDNA sequenc
18	287.5	54.0	2427	20	AAO7369	Human p2Y11 recept
19	286	53.8	2010	24	ABA02283	Human DNA replicat
20	285.5	53.7	2501	23	AAO84048	DNA encoding novel
21	285	53.6	2930	23	ABL29672	Drosophila melanog
22	285	53.6	3131	23	ABL29698	Drosophila melanog
23	284.5	53.5	2008	22	AAH16499	Human cDNA sequenc
24	284.5	53.5	2410	24	ABL56707	Nucleotide sequenc
25	284.5	53.5	2727	22	AAH14454	Human cDNA sequenc
26	284.5	53.5	3243	23	ABK42289	Genomic sequence #
27	284.5	53.5	3436	20	AAZ09231	Murine IL-10 recep
28	284.5	53.5	4594	22	AAK81445	Human immune/haema
29	284	53.4	1110	22	AAK77911	Human immune/haema
30	284	53.4	3520	15	AAQ69215	Interleukin-10 rec
31	283.5	53.3	3162	22	AAO36572	Human musculoskele
32	283	53.2	2533	19	AAV96031	Mouse gamma II ada
33	283	53.2	3520	19	AAV13207	Mouse IL-10 recept
34	282.5	53.1	1409	23	ABV24915	Human prostate exp
35	282.5	53.1	2342	12	AAQ14955	Rat dopamine D1 re
36	282.5	53.1	3025	12	AAQ13337	D1 dopamine recept
37	282.5	53.1	3025	18	AAQ63657	D1 dopamine recept
38	282	53.0	1001	21	AAC57685	Arachidonic acid m
39	282	53.0	3127	23	ABL03828	Drosophila melanog
40	281.5	52.9	1635	23	AAS69998	DNA encoding novel
41	281.5	52.9	4344	24	ABN95256	Gene #1754 used to
42	281.5	52.9	4388	18	AAV75302	Nucleotide sequenc
43	281	52.8	2693	14	AAQ38220	NANBH virus strain
44	281	52.8	4168	21	AAZ93359	Sequence encoding
45	280.5	52.7	1807	22	AAK87100	Human immune/haema

ALIGNMENTS

RESULT 1  
AAD23462  
ID AAD23462 standard; cDNA; 337 BP.  
XX  
AC AAD23462;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE Human lung tumour-specific 20E10 5' cDNA.  
XX  
KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
KW antisense-therapy; vaccine; immune response; lung cancer; 20E10; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200172295-A2.  
XX  
PD 04-OCT-2001.

XX PF 28-MAR-2001; 2001WO-US09991.  
XX PR 29-MAR-2000; 2000US-0538037.  
PR 05-JUN-2000; 2000US-0588937.  
PR 18-AUG-2000; 2000US-0640878.  
PR 22-SEP-2000; 2000US-234517P.  
PR 01-NOV-2000; 2000US-0704512.  
PR 14-DEC-2000; 2000US-0738973.  
XX PA (CORI-) CORIXA CORP.  
XX PI Reed SG, Lodes MJ, Mohamath R, Secrlist H, Benson DR, Indirias CY;  
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
XX DR WPI; 2001-639201/73.  
XX PT New human lung-specific polynucleotides and polypeptides for the  
PT diagnosis and treatment of disease e.g. lung cancer -  
XX PS Claim 1; Page 334; 378pp; English.  
XX CC The invention relates to isolated lung tumour-specific proteins and  
CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
CC their antigen-presenting cells are useful for stimulating and/or  
CC expanding T cells specific for a tumour protein, and for inhibiting  
CC the development of cancer. The invention also relates to a composition  
CC useful for stimulating an immune response, and for treating cancer. The  
CC lung tumour specific oligonucleotide is useful in gene therapy and for  
CC diagnosis, detection and treatment of lung cancer. The present sequence  
CC is a cDNA encoding human lung tumour-specific protein.  
XX SQ Sequence 337 BP; 103 A; 60 C; 93 G; 81 T; 0 other;  
Alignment Scores:  
Pred. No.: 1.4e-05 Length: 337  
Score: 532.00 Matches: 97  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 23 Gaps: 0  
US-09-854-133-586 (1-97) x AAD23462 (1-337)  
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
Db 5 GAGGTTGAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTGAAGTCTGTCTCAACA 64  
QY 21 GluLeuArgLysLysGluArgLysLysLysLysLysLysPheGlnAlaAsnCysGly 40  
Db 65 GAATTAAGGAAAGAAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 124  
QY 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisTrpIleGln 60  
Db 125 ATAGATTTTATCATATCTGGATTTTGGATTTCTTTTCTCTCATCTGGATTTCAG 184  
QY 61 GluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
Db 185 GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACCTGCAGGAAATGTTAACGGGA 244  
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97  
Db 245 GGCTGCCTTCCCTGGGCAACAAGGAGGCCACCTGGGAGGAGAAAGTGCAGC 295  
RESULT 2  
AAD23460  
ID AAD23460 standard; cDNA; 2239 BP.  
XX AC AAD23460;  
XX DT 26-FEB-2002 (first entry)  
XX DE Human lung tumour-specific 19A4 cDNA.

XX KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
KW antisense-therapy; vaccine; immune response; lung cancer; 19A4; ss.  
XX OS Homo sapiens.  
XX PN WO200172295-A2.  
XX PD 04-OCT-2001.  
XX PF 28-MAR-2001; 2001WO-US09991.  
XX PR 29-MAR-2000; 2000US-0538037.  
PR 05-JUN-2000; 2000US-0588937.  
PR 18-AUG-2000; 2000US-0640878.  
PR 22-SEP-2000; 2000US-234517P.  
PR 01-NOV-2000; 2000US-0704512.  
PR 14-DEC-2000; 2000US-0738973.  
XX PA (CORI-) CORIXA CORP.  
XX PI Reed SG, Lodes MJ, Mohamath R, Secrlist H, Benson DR, Indirias CY;  
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
XX DR WPI; 2001-639201/73.  
XX PT New human lung-specific polynucleotides and polypeptides for the  
PT diagnosis and treatment of disease e.g. lung cancer -  
XX PS Claim 1; Page 332; 378pp; English.  
XX CC The invention relates to isolated lung tumour-specific proteins and  
CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
CC their antigen-presenting cells are useful for stimulating and/or  
CC expanding T cells specific for a tumour protein, and for inhibiting  
CC the development of cancer. The invention also relates to a composition  
CC useful for stimulating an immune response, and for treating cancer. The  
CC lung tumour specific oligonucleotide is useful in gene therapy and for  
CC diagnosis, detection and treatment of lung cancer. The present sequence  
CC is a cDNA encoding human lung tumour-specific protein.  
XX SQ Sequence 2239 BP; 619 A; 444 C; 493 G; 683 T; 0 other;  
Alignment Scores:  
Pred. No.: 0.0127 Length: 2239  
Score: 523.00 Matches: 97  
Percent Similarity: 84.35% Conservative: 0  
Best Local Similarity: 84.35% Mismatches: 0  
Query Match: 98.31% Indels: 18  
DB: 23 Gaps: 1  
US-09-854-133-586 (1-97) x AAD23460 (1-2239)  
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
Db 2 GAGGTTGAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTGAAGTCTGTCTCAACA 61  
QY 21 GluLeuArgLysLysGluArgLysLysLysLysLysLysPheGlnAlaAsnCysGly 40  
Db 62 GAATTAAGGAAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121  
QY 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisTrpIleGln 60  
Db 122 ATAGATTTTATCATATCTGGATTTTGGATTTCTTTTCTCTCATCTGGATTTCAG 181  
QY 61 GluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
Db 182 GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACCTGCAGGAAATGTTAACGGGA 241  
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArg----- 93  
Db 242 GGCTGCCTTCCCTGGGCAACAAGGAGGCCACCTGGGAGGAGAGAGAGAGAGAGAGAG 301

Pred. No.: 103 Length: 800  
Score: 339.50 Matches: 62  
Percent Similarity: 57.25% Conservative: 13  
Best Local Similarity: 47.33% Mismatches: 7  
Query Match: 63.82% Indels: 49  
DB: 20 Gaps: 28

US-09-854-133-586 (1-97) x AAZ16609 (1-800)

QY 1 GluValGluValSerArg---AspHis---Ala-----Ser-----Leu 11  
DB 107 GAG---GAGGTGGAGAAATTGAGAGACACGATGCATACACAGGTGTTCTGAGTAGTAATTA 163  
QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22  
DB 164 ---GATCGCTGTGAAGGAAAGACACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 217  
QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33  
DB 218 CGCTGAGAGAGACAGTCTGAAAGCAGAGAGAGACATCGATCAGTAACACCAAGAGACACC 277  
QY 34 LysPheGlnAlaAsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---Phe 49  
DB 278 AAA-----GTTGAAAGTTT---GTTTCTTCTCCCTCTGTTTATT 316  
QY 50 Trp-----Ile---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysPro 66  
DB 317 TTCCCCCGGTGTCTCCCTACTA-----TGG---TCAGAAAGCCTGTTGTGTCCA 361  
QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla 86  
DB 362 CCATCTCCAAAGGAGGTACCTGCAGGGAATGTAACGGGAGGCTGCCTTCCCTGGGCA 421

QY 87 ThrArgSerHisLeuGlyArgArgLysCysSer 97  
DB 422 ACAAGGAGCCACCTGGCAGGAGAAAGTGCAGC 454

RESULT 4  
ABL37718  
ID ABL37718 standard; cDNA; 575 BP.  
XX AC ABL37718;  
XX DT 08-APR-2002 (first entry)  
XX DE Human colon tumour antigen polynucleotide SEQ ID NO:1307.  
XX KW Human; colon cancer; colon tumour antigen; cytostatic; vaccine;  
XX KW colon tumour metastatic antigen; diagnosis; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200196388-A2.  
XX PD 20-DEC-2001.  
XX PF 08-JUN-2001; 2001WO-US18557.  
XX PR 09-JUN-2000; 2000US-210899P.  
XX PR 20-FEB-2001; 2001US-270216P.  
XX PA (CORI-) CORIXA CORP.  
XX PI Jiang Y, Harlocker SL, Secrist H;  
XX WPI; 2002-114514/15.  
XX DR Novel isolated colon tumor polynucleotide differentially expressed in  
XX PT colon tumor or colon metastatic tumor and polypeptides encoded by them,  
XX PT useful for inhibiting development of cancer in patient -  
XX PS Claim 1; SEQ ID 1307; 105pp; English.

QY 94 -----ArgLysCysSer 97  
DB 302 CCTTTTCAGGAGACAGCGCTTTTCAGGAGAGAGAAAGTGCAGC 346

RESULT 3  
AAZ16609  
ID AAZ16609 standard; cDNA; 800 BP.  
XX AC AAZ16609;  
XX DT 12-OCT-1999 (first entry)  
XX DE Human gene expression product cDNA sequence SEQ ID NO:4079.  
XX KW Human; gene; gene expression product; diagnosis; therapy; probe;  
KW detection; mapping; tissue typing; profiling; forensic; cancer;  
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
XX OS Homo sapiens.  
XX PN WO9938972-A2.  
XX PD 05-AUG-1999.  
XX PF 28-JAN-1999; 99WO-US01619.  
XX PR 03-APR-1998; 98US-0080666.  
XX PR 28-JAN-1998; 98US-0072910.  
XX PR 24-FEB-1998; 98US-0075954.  
XX PR 31-MAR-1998; 98US-0080114.  
XX PR 03-APR-1998; 98US-0080515.

XX PA (CHIR ) CHIRON CORP.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;  
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;  
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;  
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;  
XX WPI; 1999-494092/41.  
XX DT Novel human genes and their expression products which are  
XX PT differentially expressed in different cell types  
XX PS Claim 1; Page 1934; 2479pp; English.

XX CC The present invention describes a library of human polynucleotides  
XX CC comprising the sequences given in AAZ12532 to AAZ1779. Also described is  
XX CC a method of detecting differentially expressed genes correlated with the  
XX CC cancerous state of a mammalian cell, comprising detecting at least one  
XX CC differentially expressed gene product in a test sample from a cell  
XX CC suspected of being cancerous, where the gene product is encoded by one  
XX CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The  
XX CC polynucleotides can be used as a source of primers and probes, which can  
XX CC be used for a variety of purpose, e.g. detection of expression levels,  
XX CC mapping, tissue typing or profiling, forensics, genetic analysis and  
XX CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
XX CC can be used for raising antibodies for experimental, diagnostic and  
XX CC therapeutic purposes. The polynucleotides may also be used to construct  
XX CC arrays for diagnostics (which may be used to determine function of an  
XX CC encoded protein); and to detect differences in expression levels between  
XX CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
XX CC identify a genetic predisposition or susceptibility to a disease such as  
XX CC cancer). The polynucleotides of the invention are especially used in the  
XX CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
XX CC and lung cancer. The polynucleotides can also be used to screen for  
XX CC peptide analogues and antagonists.

XX SQ Sequence 800 BP; 199 A; 166 C; 202 G; 211 T; 22 other;  
Alignment Scores:



ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I) which were isolated from human colon tumour and colon metastatic tumour cDNA libraries. (I) have cytostatic activity and can be used in vaccine production. (I) can be used for stimulating and/or expanding T cells specific for a tumour protein on contact with the T cells. They are also useful for inhibiting the development of cancer in a patient. (I) can be used as probes or primers for nucleic acid hybridisation, for preparing mutant species primers, or primers for use in genetic constructions. (I) can be used in the diagnosis of a colon tumour.

SQ Sequence 575 BP; 151 A; 122 C; 153 G; 141 T; 8 other;

**Alignment Scores:**

Pred. No.:	36.2	Length:	575
Score:	339.00	Matches:	62
Percent Similarity:	57.25%	Conservative:	13
Best Local Similarity:	47.33%	Mismatches:	6
Query Match:	63.72%	Indels:	50
DB:	24	Gaps:	28

US-09-854-133-586 (1-97) x ABL37718 (1-575)

QY	1	GluValGluValSerArg	---AspHis---Ala---	---Ser---	Leu	11
			:::			
Db	22	GAGGTGGAG	-----AATTGAGAGCACGATGCATACACAGGTGTTTCTGAGTAGTAATA		75	
QY	12	GlyAsp	---SerGlu---	Thr---LeuSer---	ThrGluLeu---	22
Db	76	---GATCGCTGTGAAGGAAAAGCACACCTTTGAGTTTTCACCTGTGAACA		---CTATAG	129	
QY	23	Arg	-----LysLys---	Glu---Arg---	Lys---	33
			:::		:::	
Db	130	CGCTGAGAGAGACAGTCTGAAAGCAGAGGAGACATCGATCAGTAACACCCAGACAGACACC			189	
QY	34	LysPheGlnAlaAsnCysGlyIleAsp	---PheIleIlePheTrp---	Ile---	Phe	49
			:::		:::	
Db	190	AAA	-----GTTGAAAGTTT---	GTTTCTTCCCTCTGTTTATT	228	
QY	50	Trp	-----Ile---	LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysPro	66	
		:::				
Db	229	TTCCCCCGTGTGTCCTACTA	-----TGG---	TCAGAAAGCCTGTTGTGTCCA	273	
QY	67	ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla			86	
Db	274	CCATCTCCAAAGGAGGTTACCTGCAGGGAATGTTAACGGGAGGCTGCCTTCCCTGGGCA			333	
QY	87	ThrArgSerHisLeuGlyArgArgLysCysSer			97	
Db	334	ACAAGGAGCCACCTGGGCAGGAGAAAGTGCAGC			366	

## RESULT 5

ABK27651/c  
ID ABK27651 standard; cDNA; 596 BP.

AA  
AC

XX  
DT  
vv  
09-APR-2002 (first entry)

Accession	Human colon cancer expressed sequence tag, Seq ID no 88.
AA010000	
DE010000	
XX010000	
KW010000	Human; colon cancer; T cell expansion; tumour; EST; gene expressed sequence tag.
KW010000	

XX  
OS  
XX  
PN

XX  
PD  
20-DEC-2001

08-JUN-2001; 2001WO-US18577.

PR 09-JUN-2000; 2000US-210821P.  
PR 18-DEC-2000; 2000US-256571P.

10-MAY-2001; 2001US-290240P.  
(CORI-) CORIXA CORP.

Jiang Y, Hepler WT, Clapper JD, Wang A, Secretist H;  
WPI: 2002-139708/18.

Novel isolated polynucleotide encoding a polypeptide comprising a portion of colon tumour protein, useful for detection, diagnosis and therapy of human colon cancer -

Claim 1; Page 174; 220pp; English.

The invention relates to an isolated polynucleotide (I) encoding a polypeptide (II) comprising at least a portion of a colon tumour protein. (I), (II) and antibody (III) to (II) are useful for determining the presence of a cancer in a patient. (I), (II) or antigen presenting cells expressing (I) is useful for stimulating and/or expanding T cells specific for a tumour protein, by contacting T cells with (I), (II) or antigen-presenting cells that express (I), under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells. (I), (II), or antigen presenting cells that express (II) are useful for treating colon cancer in a patient by incubating CD4+ and/or CD8+ T cells isolated from a patient with (I), (II) or antigen presenting cells that express (II), such that T cells proliferate, and administering to the patient an effective amount of the proliferated T cells, thus inhibiting the development of a cancer in the patient. (I) or (II) is useful in vaccines and pharmaceutical compositions for prevention and treatment of colon malignancies and for the diagnosis and monitoring of such cancers. (I), (II) or (III) is useful for detection, diagnosis and/or therapy of human colon cancer. (I) is useful as a probe or primer for nucleic acid hybridisation, and in the design and preparation of ribozyme molecules for inhibiting expression of (II) in tumour cells. ABK27564-ABK27807 represent novel human colon cancer coding sequences and primers of the invention.

SQ Sequence 596 BP; 160 A; 151 C; 126 G; 158 T; 1 other;

**Alignment Scores:**

Pred. No.:	88.3	Length:	596
Score:	327.50	Matches:	59
Percent Similarity:	57.48%	Conservative:	14
Best Local Similarity:	46.46%	Mismatches:	5
Query Match:	61.56%	Indels:	49
DB:	24	Gaps:	27

US-09-854-133-586 (1-97) x ABK27651 (1-596)

QY	6	Arg	-----AspHis	---Ala	-----Ser	-----LeuGlyAsp	---Ser	14
		:::	:::					
Db	595	AAGAATTGAGAGC	CAGATGCATACACAGG	TGTTTCTGAGTAGTA	ATTA	---GATCGCTGT		5399
QY	15	Glu	-----Thr	---LeuSer	---Gln	-----ThrGluLeu	---Arg	Lys
								24
Db	538	GAAGNAAAAAGC	ACACCTTTGAGTTT	TACCTGTGAACA	---CTATAGCGCTGAGAGAGA		:::	482
QY	25	Lys	---Glu	---Arg	---Lys	---LysLys	---ArgGluArgLysPheGlnAla	37
		:::		:::	:::	:::		
Db	481	CAGTCTGAAGCAG	AGGAACATCGATCAGTAACACCAAGAGACACCAAA					431
QY	38	AsnCysGlyIleAsp	---PheIleIlePheTrp	-----Ile	---PheTrp	-----Ile	51	
		:::>:::		:::>:::>:::	:::	:::	:::	
Db	430	-----GTTGAAAGTTT	---GTTTTCCTCCCTCTGTTTATTTT	TTCCTCCCGGTGTG			383	
QY	52	---LeuLeuPheSerHis	HisTrpIleGlnGluSerLeuLeuCysProProSerProLys				70	
Db	382	TCCCTACTA	-----TGG	---TCAGAAAGCCTGTTGTGTCCACCATCTCCAAAG			338	
QY	71	GluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAlaThrArgSerHis					90	
Db	337	GAGGTTACCTGCAGGGAAATGTTAACGGGAGGCTGCCTTCCCTGGGCAACAAGGAGGCCAC					278	



us-09-854-133-586.rng

Tue May 13 12:12:49 2003

```
QY      91 LeuGlyArgArgLysCysSer 97
      |||||
Db      277 CTGGCAGGAGAGAAAGTGCAGC 257

RESULT 6
ABAL6566/c
ID      ABAL6566 standard; DNA; 2752 BP.
XX
AC      ABAL6566;
XX
DT      23-JAN-2002 (first entry)
XX
DE      Human nervous system related polynucleotide SEQ ID NO 8897.
XX
KW      Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW      immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW      antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW      antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW      antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW      antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW      neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS      Homo sapiens.
XX
PN      WO200159063-A2.
XX
PD      16-AUG-2001.
XX
PF      17-JAN-2001; 2001WO-US01334.
XX
PR      31-JAN-2000; 2000US-0179065.
PR      04-FEB-2000; 2000US-0180628.
PR      24-FEB-2000; 2000US-0184664.
PR      02-MAR-2000; 2000US-0186350.
PR      16-MAR-2000; 2000US-0189874.
PR      17-MAR-2000; 2000US-0190076.
PR      18-APR-2000; 2000US-0198123.
PR      19-MAY-2000; 2000US-0205515.
PR      07-JUN-2000; 2000US-0209467.
PR      28-JUN-2000; 2000US-0214886.
PR      30-JUN-2000; 2000US-0215135.
PR      07-JUL-2000; 2000US-0216647.
PR      07-JUL-2000; 2000US-0216880.
PR      11-JUL-2000; 2000US-0217487.
PR      11-JUL-2000; 2000US-0217496.
PR      14-JUL-2000; 2000US-0218290.
PR      26-JUL-2000; 2000US-0220963.
PR      26-JUL-2000; 2000US-0220964.
PR      14-AUG-2000; 2000US-0224518.
PR      14-AUG-2000; 2000US-0224519.
PR      14-AUG-2000; 2000US-0225213.
PR      14-AUG-2000; 2000US-0225214.
PR      14-AUG-2000; 2000US-0225266.
PR      14-AUG-2000; 2000US-0225267.
PR      14-AUG-2000; 2000US-0225268.
PR      14-AUG-2000; 2000US-0225270.
PR      14-AUG-2000; 2000US-0225447.
PR      14-AUG-2000; 2000US-0225757.
PR      14-AUG-2000; 2000US-0225758.
PR      14-AUG-2000; 2000US-0225759.
PR      18-AUG-2000; 2000US-0226279.
PR      22-AUG-2000; 2000US-0226681.
PR      22-AUG-2000; 2000US-0226868.
PR      22-AUG-2000; 2000US-0227182.
PR      23-AUG-2000; 2000US-0227009.
PR      30-AUG-2000; 2000US-0228924.
PR      01-SEP-2000; 2000US-0229287.
PR      01-SEP-2000; 2000US-0229343.
PR      01-SEP-2000; 2000US-0229344.
PR      01-SEP-2000; 2000US-0229345.
PR      05-SEP-2000; 2000US-0229509.
PR      05-SEP-2000; 2000US-0229513.
PR      06-SEP-2000; 2000US-0230437.
PR      06-SEP-2000; 2000US-0230438.
PR      08-SEP-2000; 2000US-0231242.
PR      08-SEP-2000; 2000US-0231243.
PR      08-SEP-2000; 2000US-0231244.
PR      08-SEP-2000; 2000US-0231244.
PR      08-SEP-2000; 2000US-0231413.
PR      08-SEP-2000; 2000US-0231414.
PR      08-SEP-2000; 2000US-0232080.
PR      08-SEP-2000; 2000US-0232081.
PR      12-SEP-2000; 2000US-0231968.
PR      14-SEP-2000; 2000US-0232397.
PR      14-SEP-2000; 2000US-0232398.
PR      14-SEP-2000; 2000US-0232399.
PR      14-SEP-2000; 2000US-0232400.
PR      14-SEP-2000; 2000US-0232401.
PR      14-SEP-2000; 2000US-0233063.
PR      14-SEP-2000; 2000US-0233064.
PR      14-SEP-2000; 2000US-0233065.
PR      21-SEP-2000; 2000US-0234223.
PR      21-SEP-2000; 2000US-0234274.
PR      25-SEP-2000; 2000US-0234997.
PR      25-SEP-2000; 2000US-0234998.
PR      26-SEP-2000; 2000US-0235484.
PR      27-SEP-2000; 2000US-0235834.
PR      27-SEP-2000; 2000US-0235836.
PR      29-SEP-2000; 2000US-0236327.
PR      29-SEP-2000; 2000US-0236367.
PR      29-SEP-2000; 2000US-0236368.
PR      29-SEP-2000; 2000US-0236369.
PR      29-SEP-2000; 2000US-0236370.
PR      02-OCT-2000; 2000US-0236802.
PR      02-OCT-2000; 2000US-0237037.
PR      02-OCT-2000; 2000US-0237038.
PR      02-OCT-2000; 2000US-0237039.
PR      02-OCT-2000; 2000US-0237040.
PR      13-OCT-2000; 2000US-0239935.
PR      13-OCT-2000; 2000US-0239937.
PR      20-OCT-2000; 2000US-0240960.
PR      20-OCT-2000; 2000US-0241785.
PR      20-OCT-2000; 2000US-0241786.
PR      20-OCT-2000; 2000US-0241787.
PR      20-OCT-2000; 2000US-0241808.
PR      20-OCT-2000; 2000US-0241809.
PR      20-OCT-2000; 2000US-0241826.
PR      20-OCT-2000; 2000US-0242221.
PR      01-NOV-2000; 2000US-0244617.
PR      08-NOV-2000; 2000US-0246474.
PR      08-NOV-2000; 2000US-0246475.
PR      08-NOV-2000; 2000US-0246476.
PR      08-NOV-2000; 2000US-0246477.
PR      08-NOV-2000; 2000US-0246478.
PR      08-NOV-2000; 2000US-0246523.
PR      08-NOV-2000; 2000US-0246524.
PR      08-NOV-2000; 2000US-0246525.
PR      08-NOV-2000; 2000US-0246526.
PR      08-NOV-2000; 2000US-0246527.
PR      08-NOV-2000; 2000US-0246528.
PR      08-NOV-2000; 2000US-0246532.
PR      08-NOV-2000; 2000US-0246609.
PR      08-NOV-2000; 2000US-0246610.
PR      08-NOV-2000; 2000US-0246611.
PR      08-NOV-2000; 2000US-0246613.
PR      17-NOV-2000; 2000US-0249207.
PR      17-NOV-2000; 2000US-0249208.
PR      17-NOV-2000; 2000US-0249209.
PR      17-NOV-2000; 2000US-0249210.
PR      17-NOV-2000; 2000US-0249211.
PR      17-NOV-2000; 2000US-0249212.
PR      17-NOV-2000; 2000US-0249213.
PR      17-NOV-2000; 2000US-0249214.
PR      17-NOV-2000; 2000US-0249215.
PR      17-NOV-2000; 2000US-0249216.
PR      17-NOV-2000; 2000US-0249217.
PR      17-NOV-2000; 2000US-0249218.
```

PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-541565/60.

PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system  
PT cancers and metastases -

PS Disclosure; SEQ ID NO 8897; 1701pp + Sequence Listing; English.

XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins  
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.

CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 2752 BP; 669 A; 696 C; 666 G; 721 T; 0 other;

Alignment Scores:  
Pred. NO.: 6.41e+04 Length: 2752  
Score: 304.00 Matches: 63  
Percent Similarity: 34.82% Conservative: 23  
Best Local Similarity: 25.51% Mismatches: 6  
Query Match: 57.14% Indels: 155  
DB: 22 Gaps: 55

US-09-854-133-586 (1-97) x ABA16566 (1-2752)

QY 1 GluValGlu-----ValSerArgAspHis-----AlaSerLeu-----Gly---Asp 13  
Db 2116 GAGGTGGAGGCTGCAGTGAGCCGAGATCACACCACTGCA---CTCCAGCCTGGGCCACAG 2060  
QY 14 SerGluThrLeu---SerGlnThrGluLeuArgLysLysGluArgLysLys---Lys--- 30  
Db 2059 AGCGAGACTTGTGTTCAAAA---AAA-----AAAAA-----AAAAA-----AACT 2006  
QY 31 ---ArgGluArgLysPheGlnAla-----AsnCys---GlyIle-----Asp 42

Db 2005 GGTCAACACCGTCAATC---TCATAAATGAGCTGTGCTGGCATATGGGTGCTGTCCG 1949  
QY 43 ---PheIle-----Ile-----Phe-----Trp--- 47  
Db 1948 GCATATGTTCTAGTTATACTAAATGTCTTTCCCCAGTGGGTTTCACTCTTGTGGGCC 1889  
QY 48 Ile-----Phe-----Trp-----Ile-----LeuLeu---Phe 54  
Db 1888 ATGTGTAGTTTCTAGCTCCAGGACATTTTCAAAAGCCATTACACTAGTTATTATGTATG 1829  
QY 55 -----Ser-----His-----His 57  
Db 1828 AGGTCTAGAAAAACATGATTGAGCTGCATCCCAAGACACACGGTGCAGTTAAGTTAGCAC 1769  
QY 58 Trp-----IleGlnGluSerLeuLeu-----Cys 65  
Db 1768 TGGAGTTGATTTTCTGTGCTCCCTCCCTGCAA---AGCCTTATAACACCCCATGC 1712  
QY 66 ---Pro-Pro-----Ser-----ProLys-----Glu---Val----- 72  
Db 1711 CATCCATCCCTCTTTTCTGTACAGTCTCTGATATCCGAAGGCTTGGGACGCTATGCAAGG 1652  
QY 73 -Thr-----Cys---ArgGlu-----Met-----Leu----- 78  
Db 1651 GACCAAAATTTGTGTGTCAGCAACAAACACGAAATCGCTGATGCCTGTGACCTCACCATC 1592  
QY 79 -Thr-----Gly-----Gly-----CysLeu-----Pro-- 84  
Db 1591 GACGAGATGGAGAGTGCACCCAGCACCTGTCTTAATGGGACGTCGCTCCCGAGCCAAC 1532  
QY 85 -----TrpAla-----Thr---Arg-----SerHis----- 90  
Db 1531 GGGGATGTGGGCCCCCTCTCACACCGGACGACTATGCTGACAGGACTTTGGTCTCTGGC 1472  
QY 91 -----Leu---Gly-----Arg-----Arg---Lys----- 95  
Db 1471 TACAGCGACGAGAGCCAGACCCCTGGGAGGGATGAGGAGGACCTGGCGGATGAAATGATA 1412  
QY 96 -----CysSer 97  
Db 1411 TGCATCACCCCTTGTAGC 1393

RESULT 7  
ABA07916  
ID ABA07916 standard; DNA; 1687 BP.  
XX ABA07916;  
XX  
DT 11-JAN-2002 (first entry)  
XX  
DE Human ovarian and breast cancer associated polynucleotide SEQ ID NO 711.  
XX

KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; human; secreted protein; ds.  
OS Homo sapiens.

XX WO200155325-A2.  
PN  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01345.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.

us-09-854-133-586.rng

Tue May 13 12:12:49 2003

PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-488786/53.

New isolated ovarian and/or breast cancer related nucleic acids and polypeptides, useful for diagnosing, treating and/or preventing human diseases and disorders, particularly ovarian and/or breast cancer -

Disclosure; SEQ ID NO 711; 577pp + Sequence Listing; English.

The invention relates to novel genes (ABA07454-ABA08224) and proteins (ABB10743-ABB10980) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are

CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1687 BP; 543 A; 332 C; 356 G; 456 T; 0 other;

## Alignment Scores:

Pred. No.:	1.88e+04	Length:	1687
Score:	298.50	Matches:	58
Percent Similarity:	39.78%	Conservative:	16
Best Local Similarity:	31.18%	Mismatches:	10
Query Match:	56.11%	Indels:	102
DB:	22	Gaps:	38

US-09-854-133-586 (1-97) x ABA07916 (1-1687)

QY	1	GluValGlu-----ValSerArgAspHisAla---SerLeu-----Gly---AspSer	14
Db	633	GAGGTAGAGGTTGCAGTGAGTCGAGATCAGCCACTGCACCTTCAGCCCTGGGTGACAGAGC	692
QY	15	GluThrLeuSerGlnThrGluLeuArgLysLysGlu-----ArgLys-----Lys	29
Db	693	AAGACTTGTCTCAA-----AAAAAAAGACCTACTCTATAGAAAGCAATTTCAG	743
QY	30	LysArg---GluArgLysPheGlnAlaAsn-Cys-----Gly-----40	
Db	744	AAAAGAGTAAAC---AACTATATG---AATATGTTAGTTCAAGTAGTAATCAGGACAGTG	797
QY	41	-----Ile-----AspPheIleIle-----PheTrp-----IlePhe	49
Db	798	CAAAATCAATAACAGTGAGAT---CAAGTGTCCTCTTATTTGGTAAATAATTAACATTTT	854
QY	49	eTrpIleLeuLeu-----PheSer-----His---HisTrp-----IleG1	60
Db	855	TTTTTTCTTTTGGAGATGCCTTCTCCTTGTACCCAGGTTGGAGTGCAGTGGTGCA	914
QY	60	nGluSer---LeuLeu-----CysProProSer---68	
Db	915	A---TCTGGGCTCATTTGCAACCTCTGCCTCCAGGCTCAATCAGTCTGCCACCTCAGCC	971
QY	69	-ProLys-----GluVal-----Thr-----Cys-----74	
Db	972	TCCCAAGTAGCTGTGACTACAGGTCTCCCAACGCTGGCTAATTTTCTTGCAATTT	1031
QY	75	---ArgGluMetLeuThrGly-----GlyCysLeuPro---TrpAla---Thr---87	
Db	1032	TGATAGAGAT-----GGAGTTTCGCCAGGTTGC---CCAAGCTGGTCTCAAACTCC	1079
QY	88	-----Arg---Ser---HisLeuGly-----ArgArg-----94	
Db	1080	TGAGCTCAAGCAATCCGCCACCTTGGCTTCCCAAGTACTGGGCTTACAGGCGTGAGCC	1139
QY	95	-----Lys	95
Db	1140	ACCGGACCCAGCAAAA	1155

RESULT 8

AAL03733

ID AAL03733 standard; DNA; 1687 BP.

XX

AC AAL03733;

XX	21-NOV-2001	(first entry)
DT	Human reproductive system related antigen DNA SEQ ID NO: 6421.	
XX	Human reproductive system related antigen; reproductive system disorder;	
DE	cancer; gene therapy; ds.	
XX	Homo sapiens.	
OS	WO200155320-A2.	
PN	02-AUG-2001.	
XX	17-JAN-2001; 2001WO-US01339.	
PR	31-JAN-2000; 2000US-0179065.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0189874.	
PR	17-MAR-2000; 2000US-0190076.	
PR	18-APR-2000; 2000US-0198123.	
PR	19-MAY-2000; 2000US-0205515.	
PR	07-JUN-2000; 2000US-0209467.	
PR	28-JUN-2000; 2000US-0214886.	
PR	30-JUN-2000; 2000US-0215135.	
PR	07-JUL-2000; 2000US-0216647.	
PR	07-JUL-2000; 2000US-0216880.	
PR	11-JUL-2000; 2000US-0217487.	
PR	11-JUL-2000; 2000US-0217496.	
PR	14-JUL-2000; 2000US-0218290.	
PR	26-JUL-2000; 2000US-0220963.	
PR	26-JUL-2000; 2000US-0220964.	
PR	14-AUG-2000; 2000US-0224518.	
PR	14-AUG-2000; 2000US-0224519.	
PR	14-AUG-2000; 2000US-0225213.	
PR	14-AUG-2000; 2000US-0225214.	
PR	14-AUG-2000; 2000US-0225266.	
PR	14-AUG-2000; 2000US-0225267.	
PR	14-AUG-2000; 2000US-0225268.	
PR	14-AUG-2000; 2000US-0225270.	
PR	14-AUG-2000; 2000US-0225447.	
PR	14-AUG-2000; 2000US-0225757.	
PR	14-AUG-2000; 2000US-0225758.	
PR	14-AUG-2000; 2000US-0225759.	
PR	18-AUG-2000; 2000US-0226279.	
PR	22-AUG-2000; 2000US-0226681.	
PR	22-AUG-2000; 2000US-0226868.	
PR	22-AUG-2000; 2000US-0227182.	
PR	23-AUG-2000; 2000US-0227009.	
PR	30-AUG-2000; 2000US-0228924.	
PR	01-SEP-2000; 2000US-0229287.	
PR	01-SEP-2000; 2000US-0229343.	
PR	01-SEP-2000; 2000US-0229344.	
PR	05-SEP-2000; 2000US-0229345.	
PR	05-SEP-2000; 2000US-0229509.	
PR	06-SEP-2000; 2000US-0229513.	
PR	06-SEP-2000; 2000US-0230437.	
PR	06-SEP-2000; 2000US-0230438.	
PR	08-SEP-2000; 2000US-0231242.	
PR	08-SEP-2000; 2000US-0231243.	
PR	08-SEP-2000; 2000US-0231244.	
PR	08-SEP-2000; 2000US-0231413.	
PR	08-SEP-2000; 2000US-0231414.	
PR	08-SEP-2000; 2000US-0232080.	
PR	12-SEP-2000; 2000US-0232081.	
PR	14-SEP-2000; 2000US-0231968.	
PR	14-SEP-2000; 2000US-0232397.	
PR	14-SEP-2000; 2000US-0232398.	
PR	14-SEP-2000; 2000US-0232399.	
PR	14-SEP-2000; 2000US-0232400.	
PR	14-SEP-2000; 2000US-0232401.	



PR	14-SEP-2000;	2000US-0233063.
PR	14-SEP-2000;	2000US-0233064.
PR	14-SEP-2000;	2000US-0233065.
PR	21-SEP-2000;	2000US-0234223.
PR	21-SEP-2000;	2000US-0234274.
PR	25-SEP-2000;	2000US-0234997.
PR	25-SEP-2000;	2000US-0234998.
PR	26-SEP-2000;	2000US-0235484.
PR	27-SEP-2000;	2000US-0235834.
PR	27-SEP-2000;	2000US-0235836.
PR	29-SEP-2000;	2000US-0236327.
PR	29-SEP-2000;	2000US-0236367.
PR	29-SEP-2000;	2000US-0236368.
PR	29-SEP-2000;	2000US-0236369.
PR	29-SEP-2000;	2000US-0236370.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	02-OCT-2000;	2000US-0237040.
PR	13-OCT-2000;	2000US-0239935.
PR	13-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249267.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.

Db 1140 ACCGACCCAGCAAAA 1155  
RESULT 9  
ABA07917  
ID ABA07917 standard; DNA; 1688 BP.  
XX AC ABA07917;  
XX DT 11-JAN-2002 (first entry)  
XX DE Human ovarian and breast cancer associated polynucleotide SEQ ID NO 712.  
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antifungal; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; ds.  
OS Homo sapiens.  
XX WO200155325-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01345.  
XX PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217496.  
PR 26-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 18-AUG-2000; 2000US-0225759.  
PR 22-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 06-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 25-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 26-SEP-2000; 2000US-0234998.  
PR 27-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 29-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 02-OCT-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 20-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.

PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Barash SC, Ruben SM;

PI WPI; 2001-488786/53.

DR New isolated ovarian and/or breast cancer related nucleic acids and  
XX polypeptides, useful for diagnosing, treating and/or preventing human  
PT diseases and disorders, particularly ovarian and/or breast cancer -  
PT  
XX

PS Disclosure; SEQ ID NO 712; 577pp + Sequence Listing; English.

XX The invention relates to novel genes (ABA07454-ABA08224) and proteins  
CC (AB10743-AB10980) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.

CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1688 BP; 542 A; 332 C; 358 G; 456 T; 0 other;

Alignment Scores:  
Pred. No.: 1.89e+04 Length: 1688  
Score: 298.50 Matches: 58  
Percent Similarity: 39.78% Conservative: 16  
Best Local Similarity: 31.18% Mismatches: 10  
Query Match: 56.11% Indels: 102  
DB: 22 Gaps: 38

US-09-854-133-586 (1-97) x ABA07917 (1-1688)

QY 1 GluValGlu-----ValSerArgAspHisAla---SerLeu-----Gly---AspSer 14  
Db 634 GAGGTGGAGGTTGCAGTGCAGATCAGCCACTGCACCTCAGCTGGGTGACAGAGC 693  
QY 15 GluThrLeuSerGlnThrGluLeuArgLysLysGlu-----ArgLys-----Lys 29  
Db 694 AAGACTTTGTCTCAA-----AAAAAAGACCTACTCTATAGGAAGCAATTTCAG 744  
QY 30 LysArg---GluArgLysPheGlnAlaAsn-Cys-----Gly----- 40  
Db 745 AAAAGAGTAAAC---AACTATATG---AATATGTTAGTTTTCAGTAGTAATCAGGACAGTG 798

QY 41 ----Ile-----AspPheIleIle-----PheTrp-----IlePh 49  
Db 799 CAAATCAATAACAGTGAGAT---CAAGTGTCTCCTTATTTGGTAAAAAATAAACTTTT 855  
QY 49 eTrpIleLeuLeu-----PheSer-----His---HisTrp-----IleG1 60  
Db 856 TTTTCTCTTTTGGAGATGCCTTCTCACTTTGTCAACCCAGGTTGGAGTGCAGTGGTGCA 915  
QY 60 nGluSer---LeuLeu-----CysProProSer--- 68  
Db 916 A---TCTGGCTCATTTGCAACCTCTGCCTCCAGGCTCAATCAGTCTGCCACCTCAGCC 972  
QY 69 -ProLys-----GluVal-----Thr-----Cys----- 74  
Db 973 TCCCAAGTAGCTGTGACTACAGGTCTCTCCACCACGCTGGCTAATTTCTTGCATTTT 1032  
QY 75 ----ArgGluMetLeuThrGly-----GlyCysLeuPro---TrpAla---Thr--- 87  
Db 1033 TGATAGAGAT-----GGAGTTTCCGACGTTGC---CCAAGCTGGTCTCAAACTCC 1080  
QY 88 -----Arg---Ser---HisLeuGly-----ArgArg----- 94  
Db 1081 TGAGCTCAAGCAATCCGCCACCTTGGCTTCCAAAAGTACTGGGCTTACAGGCGTGAGCC 1140  
QY 95 -----Lys 95  
Db 1141 ACCGGACCCAGCAAAA 1156  
RESULT 10  
AAL03734  
ID AAL03734 standard; DNA; 1688 BP.  
AC AAL03734;  
XX 21-NOV-2001 (first entry)  
DE Human reproductive system related antigen DNA SEQ ID NO: 6422.  
XX Human reproductive system related antigen; reproductive system disorder;  
KW Human; reproductive system related antigen; cancer; gene therapy; ds.  
XX Homo sapiens.  
XX WO200155320-A2.  
XX 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01339.  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.

1 GluValGlu-----ValSerArgAspHisAla-----SerLeu-----Gly---AspSer 14  
 |||||  
 634 GAGGTGGAGGTTGCAGTGCAGTCGAGATCAGCCACTGCACCTCAGCTGGGTGACAGAGC 69



CC suspected of having endometrial cancer comprising determining the  
CC expression of a set of nucleic acid molecules or expression products in  
CC an endometrial sample suspected of being cancerous, where the set of  
CC nucleic acid molecules comprises at least 2 nucleic acid molecules  
CC selected from 50 fully defined sequences as given in the specification.  
CC The nucleic acids are used as an array of at least 2 of the 50  
CC nucleic acids bound to a solid substrate. Also included is a solid-phase  
CC protein microarray comprising at least 2 antibodies or its antigen  
CC binding fragments, that specifically bind at least 2 different  
CC polypeptides from the 50 fully defined sequences as given in the  
CC specification, fixed to a solid substrate. The methods and arrays are  
CC useful for the diagnosis of endometrial cancer, selecting and monitoring  
CC treatment regimes and identification of lead compounds useful for the  
CC treatment of endometrial cancer. The present sequence is one of 50  
CC genes differentially expressed between cancerous and non-cancerous  
CC samples.

XX Sequence 3995 BP; 901 A; 1113 C; 1160 G; 821 T; 0 other;  
SQ

Alignment Scores: 4.5e+05 Length: 3995  
Pred. No.: 292.50 Matches: 69  
Score: 24.53% Conservative: 9  
Percent Similarity: 21.70% Mismatches: 11  
Best Local Similarity: 54.98% Indels: 229  
Query Match: 24 Gaps: 68  
DB:

US-09-854-133-586 (1-97) x ABK35489 (1-3995)

QY 1 GluVal-----Glu-----Val---SerArg-----Asp--- 7  
Db 2708 GAGATAGATGGCGGGATTACCACTTTGTGTGTCCTCCGGGAGAAATGGAGAGGACATT 2767  
QY 8 -----His-----Ala-----Ser---Leu---Gly----- 12  
Db 2768 CAGCGGCACAAAGTTTCATTGAGCGCGGCCAGTACACAGCCACTCTATGGACCGTC 2827  
QY 13 AspSer-----Glu---Thr-----Leu-----Ser----- 18  
Db 2828 CAGTCCGTGGAGAGGTTGGCAGAGCGGGGAAGCACTGCTCTCGATGTCGGCCAAAT 2887  
QY 19 -----Gln---ThrGluLeu-----Arg---Lys 24  
Db 2888 GCCGTGGCGGGCTGCAGGGCGGCCACCTGCACCCCATCGCATCTTCATCCGCCCGC 2947  
QY 25 Lys---GluArg-----Lys---LysLysArg-----Glu-----ArgLys--- 34  
Db 2948 TCCCTGGAGAAATGTGCTAGAGATTAAACAAGCGGATCACAGAGGAGCAAGCCGCAAGCC 3007  
QY 35 Phe-----Gln-----AlaAsnCys----- 39  
Db 3008 TTCGACAGAGCCACCAAGCTGGAGCAGGAGTTTCACAGAGTGCTTCTCAGCCATCGTGGAG 3067  
QY 40 GlyIleAsp---Phe-----Ile-----Ile----- 45  
Db 3068 GGT---GACAGCTTTGAGGAGATCTACCACAAAGTGAGCGTGTCTATCGAGGACCTCTCA 3124  
QY 46 -----Phe---TrpIle-----Phe-----TrpIle---Leu 52  
Db 3125 GGCCCTTACATCTGGGTTCCAGCCCGAGAGAGACTCTGATTCCTGCTGCTGGCTGGCTG 3184  
QY 53 -----LeuPheSerHisHis-----TrpIleGlnGluSer---LeuLeu 64  
Db 3185 GACTCGCCCTGCCTC-----CATCACCTGGGCCCTTGG-----TCTGGACTGAAT 3229  
QY 65 Cys-----ProPro---Ser-----Pro----- 69  
Db 3230 TGCCCAAGCCCTTGGCTCCCGCGGCTCCCTCCACCCCTTCTTATTATTCTTCTTCT 3289  
QY 70 Lys-----Glu---Val-----Thr-----Cys-----Arg-----Glu 76  
Db 3290 AACTGGATCCAGCCTGTGTGGAGGGGGACACTCTCTGTCATGTATCCCGCAGCCAGAA 3349

QY 15 GluThrLeuSerGlnThrGluLeuArgLysLysGlu-----ArgLys-----Lys 29  
Db 694 AAGACTTTGTCTCAA-----AAAAAAGAGCCTACTCTATAGGAAAGCAATTCAG 744  
QY 30 LysArg---GluArgLysPheGlnAlaAsn-Cys-----Gly----- 40  
Db 745 AAAAGAGTAAAC---AACTATATG---AAATATGTTAGTTTCAGTAGTAATCAGGACAGTG 798  
QY 41 -----Ile-----AspPheIleIle-----PheTrp-----IlePh 49  
Db 799 CAAATCAAAATAACAGTGAGAT---CAAGTGTCTCTCTTATTGGTAAATAATTAACACTTTT 855  
QY 49 eTrpIleLeuLeu-----PheSer-----His---HisTrp-----IleG1 60  
Db 856 TTTTCTTTTGGAGATGCTCTCTCACTTTGTTCACCCAGGTTGGAGTGCAGTGGTGCA 915  
QY 60 nGluSer---LeuLeu-----CysProProSer-- 68  
Db 916 A---TCTGCGCTCATTTGCAACCTCTGCTCCCTCCAGGCTCAATCAGTCTGCCACCTCAGCC 972  
QY 69 -ProLys-----GluVal-----Thr-----Cys----- 74  
Db 973 TCCCAAGTAGCTGTGACTACAGGTCTCTCCACCGCTCCCTGCTAATTTCTTGCATTTT 1032  
QY 75 ---ArgGluMetLeuThrGly-----GlyCysLeuPro---TrpAla---Thr-- 87  
Db 1033 TGATAGAGAT-----GGAGTTTCCGCGAGTTGC---CCAAGCTGCTCTCAAACTCC 1080  
QY 88 -----Arg---HisLeuGly-----ArgArg----- 94  
Db 1081 TGAGCTCAAGCAATCCGCCACCTTGGCTTCCCAAGTACTGGGCTTACAGGCGTGAGCC 1140  
QY 95 -----Lys 95  
Db 1141 ACCGAGCCAGCAAAA 1156

RESULT 11

ABK35489  
ID ABK35489 standard; DNA; 3995 BP.

XX AC ABK35489;

XX DT 08-MAY-2002 (first entry)

XX DE Human endometrial cancer related gene, DLG4.

XX KW Human; ds; gene; endometrial cancer; differential expression;

XX KW DNA microarray; protein microarray.

XX OS Homo sapiens.

XX PN WO200209573-A2.

XX PD 07-FEB-2002.

XX PF 31-JUL-2001; 2001WO-US24104.

XX PR 31-JUL-2000; 2000US-221735P.

XX PA (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.

XX PI Mutter GL;

XX DR WPI; 2002-179967/23.

XX DR P-PSDB; AAU84269.

XX PT Diagnosing endometrial cancer comprises determining expression of  
XX PT nucleic acid molecules or expression products that are differentially  
XX PT expressed in normal and malignant endometrium -

XX PS Claim 1; Page 56-58; 233pp; English.

XX CC The invention relates to diagnosing endometrial cancer in a subject

QY 77 MetLeu-----Thr-----GlyGly-----Cys--- 82  
Db 3350 ---CTGGCTCTGAACGCCAGGACCTGGGTCTGGGGGGAGCTGGCTCTTTGTTCC 3406  
QY 83 -----Leu-----ProTrpala----- 86  
Db 3407 GAGCCCTTGCTCTTAGGATCCCC---GCCCCACCTGCCCCCAATGCACACAGACCC 3463  
QY 87 -----Thr-----ArgSer----- 89  
Db 3464 ACCGGGGGCCACCTGCCCTCCCCCATCTCTCCACACATTCAGAGTCAAGGCCCC 3523  
QY 90 -----His---Leu-----Gly---Arg-----Arg--- 94  
Db 3524 CTCGAGGAGCACCCGCTGACAGGATGCAGGGCCACAGGCTCCGCTCTCTCTTAAGGCAG 3583  
QY 95 -----Lys-----Cys-----Ser 97  
Db 3584 GGTCTGGGGTCACCCCTGCCTCATCGTAATCCCATGTTACCTTGATTCTCA 3637  
RESULT 12  
ABK35548  
ID ABK35548 standard; DNA; 3995 BP.  
XX AC ABK35548;  
AC ABK35548;  
DT 08-MAY-2002 (first entry)  
XX DE Gene DLG4 differentially expressed in breast cancer tissue.  
XX KW Human; diagnosis of breast cancer; endometrial cancer; breast tumour;  
KW MAI; mitotic activity index; cytostatic; gene; ds.  
XX OS Homo sapiens.  
XX PN WO200210436-A2.  
XX PD 07-FEB-2002.  
XX PF 27-JUL-2001; 2001WO-US23642.  
XX PR 28-JUL-2000; 2000US-222093P.  
XX PA (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.  
XX PA (BAAK/) BAAK J.  
PI Baak J, Mutter GL;  
XX WPI; 2002-180084/23.  
DR P-PSDB; AAU84328.  
XX

Diagnosing breast cancer comprises determining expression of nucleic acid molecules or expression products that are differentially expressed in normal and malignant tissue -

Claim 1; Page 58-61; 219pp; English.

The present invention relates to a method for diagnosing breast cancer in a subject suspected of having endometrial cancer. The method comprises determining the expression of a set of human genes or expression products in an endometrial sample suspected of being cancerous. The human genes of the invention are differentially expressed in breast tumours characterised as high or low MAI (mitotic activity index). These sets of genes can be used to discriminate between high and low MAI breast tumours. The invention also provides DNA and protein microarrays for analysing the expression of the human genes and their protein products. The methods and arrays are useful for the diagnosis and prognosis of endometrial cancer, selecting for the treatment regimes, and identification of compounds useful for the treatment of endometrial cancer. ABK35531-ABK35581 represent the human genes of the invention that are differentially expressed in breast cancer tissue.

SQ Sequence 3995 BP; 901 A; 1113 C; 1160 G; 821 T; 0 other;  
Alignment Scores:  
Pred. No.: 4.5e+05  
Score: 292.50  
Length: 3995  
Percent Similarity: 24.53%  
Matches: 69  
Best Local Similarity: 21.70%  
Conservative: 9  
Query Match: 54.98%  
Mismatch: 11  
Indels: 229  
DB: 24  
Gaps: 68  
US-09-854-133-586 (1-97) x ABK35548 (1-3995)  
QY 1 GluVal-----Glu-----Val---SerArg-----Asp--- 7  
Db 2708 GAGATAGATGGCCGGGATTACCACTTTGTGTCTCCCGGAGAAAATGGAGAGACATT 2767  
QY 8 -----His-----Ala-----Ser---Leu---Gly--- 12  
Db 2768 CAGCGGCACAAAGTTCAATTGAGCGCGCCAGTACACAGCCACCTCTATGGGACACGGTC 2827  
QY 13 AspSer-----Glu---Thr-----Leu-----Ser----- 18  
Db 2828 CAGTCCGTGGAGAGGTGGCAGAGCAGGGGAGACACTGCATCTCGATGTCTCGGCCAAT 2887  
QY 19 -----Gln---ThrGluLeu-----Arg---Lys 24  
Db 2888 GCGGTGGCGGGTGCAGGGCGCCACCTGCACCCCATCTCATCTCCGCCCGC 2947  
QY 25 Lys---GluArg-----Lys---LysLysArg-----Glu-----ArgLys--- 34  
Db 2948 TCCCTGAGAAATGTCTAGATATTAACAAGCGGATCACAGAGGAGCAAGCCGCAAGCC 3007  
QY 35 Phe-----Gln-----AlaAsnCys----- 39  
Db 3008 TTCGACAGAGCCACCAAGCTGGAGCAGGAGTTTCACAGAGTGTCTCAGCCATCGTGGAG 3067  
QY 40 GlyIleAsp---Phe-----Ile-----Ile----- 45  
Db 3068 GGT---GACAGCTTTGAGGAGATCTACCACAAGGTGAAGCGTGTCTATCGAGGACCTCTCA 3124  
QY 46 -----Phe---TrpIle-----Phe-----TrpIle---Leu 52  
Db 3125 GGCCCTACATCTGGTTCAGCCCGAGAGAGACTCTGTATCTCCTGCCCTGGCTGGCCTG 3184  
QY 53 -----LeuPheSerHisHis-----TrpIleGlnGluSer---LeuLeu 64  
Db 3185 GACTCGCCCTGCCTC-----CATCACCTGGCCCTTGG-----TCTGGACTGAAT 3229  
QY 65 Cys-----ProPro---Ser-----Pro----- 69  
Db 3230 TGCCCAAGCCCTTGGCTCCCGCCGCTCCCTCCACCCCTCTTATTATTCTTCT 3289  
QY 70 Lys-----Glu---Val-----Thr-----Cys-----Arg---Glu 76  
Db 3290 AACTGGATCCAGCCTGTTGGAGGGGAGACACTCTCTGCATGTATCCCGCACCCAGAA 3349  
QY 77 MetLeu-----Thr-----GlyGly-----Cys--- 82  
Db 3350 ---CTGGCTCTGAACGCCAGGAACTGGGGTCTGGGGGGAGCTGGGCTCTTGTTC 3406  
QY 83 -----Leu-----ProTrpala----- 86  
Db 3407 GAGCCCTTGCTCTTAGGATCCCC---GCCCCACCTGCCCCCAATGCACACAGACCC 3463  
QY 87 -----Thr-----ArgSer----- 89  
Db 3464 ACCGGGGGCCACCTGCCCTCCCCCATCTCTCCACACATTCAGAGTCAAGGCCCC 3523  
QY 90 -----His---Leu-----Gly---Arg-----Arg--- 94  
Db 3524 CTCGAGGAGCACCCGCTGCAGGGATGCAGGGGCCACAGGCTCCGCTCTCTCTTAAGGCAG 3583  
QY 95 -----Lys-----Cys-----Ser 97

us-09-854-133-586.rng

Tue May 13 12:12:49 2003

```
Db 3584 GGTCTGGGTACCCCTGCTCATCGTAATCCCAATGTTACCTTGATTCTCA 3637
RESULT 13
ABA07287/c
ID ABA07287 standard; DNA; 2197 BP.
XX
AC ABA07287;
XX
DT 14-JAN-2002 (first entry)
XX
DE Human pancreatic cancer related genomic DNA, SEQ ID NO: 606.
XX
KW Human; cytostatic; antidiabetic; antiinflammatory; gastric; osteopathic;
KW antihormone; antiulcer; thyroid-active; gene therapy; antisense therapy;
KW pancreatic cancer antigen inhibitor; pancreatic cancer; pancreatitis;
KW diabetes; endocrine disorder; acromegaly; hyperthyroidism;
KW gastrointestinal disorder; Crohn's disease; duodenal ulcer; ds.
XX
OS Homo sapiens.
XX
PN WO200155206-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01353.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 22-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
```

PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0251989.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0254097.  
XX 05-JAN-2001; 2001US-0259678.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-457717/49.  
XX  
PT Isolated pancreatic cancer polypeptide for treating, preventing and/or  
PT prognosing disorders related to the pancreas including pancreatic  
PT cancers and also for testing and detection e.g. diagnosis -  
XX Disclosure; SEQ ID NO 606; 537pp; English.  
PS  
XX The invention relates to an isolated polypeptide comprising an amino  
CC acid sequence at least 90% identical to 188 amino acid sequences fully  
CC defined in the specification and encoded by 188 cDNA clones fully  
CC having biological activity, a domain, an epitope, full length protein,  
CC variant, allelic variant or a species homologue of the fully defined  
CC sequence. The polynucleotide and polypeptide are useful for treating,  
CC preventing and/or prognosing disorders related to the pancreas including  
CC pancreatic cancer, pancreatitis, diabetes, endocrine disorders such as  
CC acromegaly or hyperthyroidism, and gastrointestinal disorders such as  
CC Crohn's disease and duodenal ulcers. The present sequence encodes a  
XX pancreatic cancer-related polypeptide of the invention.  
SQ Sequence 2197 BP; 493 A; 500 C; 602 G; 602 T; 0 other;  
Alignment Scores:  
Pred. No.: 6.88e+04 Length: 2197  
Score: 292.00 Matches: 59  
Percent Similarity: 34.22% Conservative: 18  
Best Local Similarity: 26.22% Mismatches: 12  
Query Match: 54.89% Indels: 136  
DB: 22 Gaps: 49  
US-09-854-133-586 (1-97) x ABA07287 (1-2197)  
QY 1 GluValGlu-----ValSerArgAspHis-----AlaSerLeu-----Gly-----Asp 13  
Db 2012 GAGCGGGAGGTTGCAGTGAGCGGAGATCACACCATTTGCA---CTCCAGCCTGGGCAACAG 1956  
QY 14 SerGluThrLeu-----SerGlnThrGluLeuArgLysLysGluArg---LysLys--- 29  
Db 1955 AGTGAGACTCTGCCTTAAAGAAAA---AAA---AAAAAAAGAGCTGACAAAGAGACA 1902  
QY 30 Lys-----Arg-----GluArg-----Lys---Phe-----GlnAla 37  
Db 1901 GAAGTGACAGGAAGTAGAGTGGACTTGAAGAGGAGATATAATGCATATAAGACAGGCC 1842  
QY 38 ---Asn-----Cys---GlyIleAsp---PheIle----- 44  
Db 1841 CTAATGCTAAGAAATTAGGACAGGCTGCTCCAGGAAGTACAGATGGCTGTGCTGG 1782  
QY 45 ---IlePhe-----Triple 48  
DB: 48  
PR 1781 TCCCTCTGGAGCAGACACTGAACAGGGCAGACCAAGTTCTCACCGAAGAAGTGGTT 1722  
QY 49 -----PheTrpIleLeuLeuPheSerHisHis-----TrpIleGln---Glu 61  
Db 1721 CCATGTTTCTGG-----CCATTTCCTCCATGGAGTGTGCACCTGG---GAGGCTGAT 1674  
QY 62 Ser-----Leu---LeuCys-----ProProSerProLysGlu-----Val 72  
Db 1673 TCTTGGACCTGGGGGTTTCTGGGTGCTGACGCGAGCCAGGAGGCGAGAGCTCAGTG 1614  
QY 73 ---ThrCys-----ArgGluMet-----LeuThr---Gly---Gly--- 81  
Db 1613 AGCACTTGCTGGTGAGAGCACTGAAACCAAGTCTCAGCCAGGAGGAGGAGGAGGAGG 1554  
QY 82 -----Cys-----Leu---Pro-----Trp--- 85  
Db 1553 CAGGCCAGAACCCACAGCCGATGCCCTCTCTGCTGCCACCACTTCTCTGTGGTCT 1494  
QY 86 -----AlaThr-----Arg-----Ser----- 89  
Db 1493 CCGTCACCTCTACGCACCTCAGGCTGCAGAGCAGCAAGCTGACCTCAGGCTAGACCCT 1434  
QY 90 -----His-----Leu---Gly---Arg 93  
Db 1433 GCAGGGAAATGCTCTTCCAAGATCATGACCTCCTATCGATAAATTGAACGGGCCCAAA 1374  
QY 94 ArgLys---CysSer 97  
Db 1373 ---AAACAAGCATCC 1362  
RESULT 14  
ID AAK89932/c  
XX AAK89932 standard; DNA; 2197 BP.  
AC AAK89932;  
XX  
DT 05-NOV-2001 (first entry)  
DE Human digestive system antigen genomic sequence SEQ ID NO: 3508.  
XX Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum; ds.  
OS Homo sapiens.  
XX  
PN WO200155314-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01324.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.



us-09-854-133-586.rng

Tue, May 13 12:12:49 2003

PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-502630/55.  
Polynucleotides encoding digestive system antigens, useful for  
diagnosing, treating, preventing and/or prognosing disorders of the  
digestive system, particularly cancer and cancer metastases -  
Disclosure; SEQ ID NO 3508; 986pp; English.  
The present invention provides the protein and coding sequences of a  
number of human digestive system antigens. These can be used in the  
diagnosis, treatment and prevention of digestive system disorders,  
including cancer, Meckel's diverticulum, bacterial or parasitic  
infections, appendicitis, Hirschsprung's disease, chronic colitis or  
ulcerative colitis. The present sequence is a genomic DNA fragment  
encoding a digestive system antigen of the invention.  
XX  
PS  
XX  
CC  
CC  
CC  
CC  
CC  
CC  
XX  
SQ  
Sequence 2197 BP; 493 A; 500 C; 602 G; 602 T; 0 other;  
Alignment Scores:  
Pred. No.: 6.88e+04  
Score: 292.00  
Percent Similarity: 34.22%  
Best Local Similarity: 26.22%  
Query Match: 54.89%  
DB: 22  
Length: 2197  
Matches: 59  
Conservative: 18  
Mismatch: 12  
Indels: 136  
Gaps: 49

PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.

US-09-854-133-586 (1-97) x AAK89932 (1-2197)

QY 1 GluValGlu-----ValSerArgAspHis-----AlaSerLeu-----Gly----Asp 13  
Db 2012 GAGCGGAGGTTGTCAGTGCAGCCGAGATCACACCATTCGA---CTCCAGCCTGGGCAACAG 1956  
QY 14 SerGluThrLeu-----SerGlnThrGluLeuArgLysLysGluArg---LysLys--- 29  
Db 1955 AGTGAGACTCTGCCTTAAAAAAA---AAA-----AAAAAAAAGAGCTGACAGAGACA 1902  
QY 30 Lys-----Arg-----GluArg-----Lys---Phe-----GlnAla 37  
Db 1901 GAACTGACAGGAGTAGAGTGGACTTGAAAAAGGAGATATAATGCATATAAGACAGGCC 1842  
QY 38 ---Asn-----Cys---GlyIleAsp---PheIle----- 44  
Db 1841 CTAATGGCTAAGAATTAGGGACAGGCTGTCTCCAGGAAGTGACAGATGGCTGTGCTGG 1782  
QY 45 ---IlePhe-----Cys-----TrpIle 48  
Db 1781 TCCCTCTGGGAGCAGACACTGAACAGGGCAGACCAAGGTTCTCACCGAAGAGTGGGTT 1722  
QY 49 -----PheTrpIleLeuLeuPheSerHis-----TrpIleGln---Glu 61  
Db 1721 CCATGTTCTGG-----CCATTTCC---CATGGAGTGTGCACCTGG---GAGGCTGAT 1674  
QY 62 Ser-----Leu---LeuCys-----ProProSerProLysGlu-----Val 72  
Db 1673 TCTTGGACCTGGGGGTTGCTGGGTGCCTGACGACCCACGAGGGCAGAGCTCAGTG 1614  
QY 73 ---ThrCys-----ArgGluMet-----LeuThr---Gly-----Gly--- 81  
Db 1613 AGCACTTGCTGGTGGAGAGCACTGAAACCAAGTCTCAGCCACGAGCCGAGGATGGGCC 1554  
QY 82 -----Cys-----Leu---Pro-----Trp--- 85  
Db 1553 CAGGCCAGAACCCACAGCCGATGCCCTCTCCTGCTGCCACACACTCCTCTGTTGCTCT 1494  
QY 86 -----AlaThr-----Arg-----Ser----- 89  
Db 1493 CCGTCACCTCTACGCACTCAGGCTGCAGAGCAGCAAGCTGACCTCAGGCTAGACCT 1434  
QY 90 -----His-----Leu---Gly---Arg 93  
Db 1433 GCAGGAAATGCTCTTCCAAGATCATGACCTCTAATCGATAAATGAACGGGCCCAA 1374  
QY 94 ArgLys---CysSer 97  
Db 1373 ---AAACAAGCATCC 1362

RESULT 15  
ABAI6568/C  
ID ABAI6568 standard; DNA; 2751 BP.  
XX AC ABAI6568;  
XX DT 23-JAN-2002 (first entry)  
XX DE Human nervous system related polynucleotide SEQ ID NO 8899.  
XX KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antitickling; antianaemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
OS Homo sapiens.  
XX PN WO200159063-A2.  
XX PD 16-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01334.  
XX PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.

us-09-854-133-586.rng

Tue May 13 12:12:49 2003

PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system  
PT cancers and metastases -  
XX Disclosure; SEQ ID NO 8899; 1701pp + Sequence Listing; English.  
PS  
XX The invention relates to novel genes (AB11004-ABA21534) and proteins  
CC (AB114678-AB18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, mellitus, Crohn's  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus and ulcerative  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2751 BP; 668 A; 696 C; 665 G; 722 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 1.52e+05 Length: 2751  
Score: 291.00 Matches: 63  
Percent Similarity: 34.82% Conservative: 23  
Best Local Similarity: 25.51% Mismatches: 6  
Query Match: 54.70% Indels: 156  
DB: 22 Gaps: 55  
  
US-09-854-133-586 (1-97) x ABA16568 (1-2751)  
QY 1 GluValGlu-----ValSerArgAspHis-----AlaSerLeu-----Gly---Asp 13  
Db 2115 GAGGTGGAGGCTGCAGTGCAGCGAGATCACACCACTGCA---CTCCAGCCTGGGCCACAG 2059  
QY 14 SerGluThrLeu---SerGlnThrGluLeuArgLysLysGluArgLysLys---LysArg 31  
Db 2058 AGCGAGACTCTGTCTCAAAA---AAA-----AAAAAAAAAAAAAAAAAAAAAACT- 2006  
QY 32 -----GluArgLysPheGlnAla-----AsnCys---GlyIle-----Asp 42  
Db 2005 GGTACACCACTGCAATTC---TCATAAATGAGCTGTGCTGGCATTATGGTGTGTTCCG 1949  
QY 43 ---PheIle-----Ile-----Phe-----Trp--- 47  
Db 1948 GCATATGTTTCTAGTTTATACTAAATGTCTTCCCCAGTGGTTCCTCTGTGGGCC 1889  
QY 48 Ile-----Phe-----Trp-----Ile-----LeuLeu---Phe 54  
Db 1888 ATGTGTAGTGTCTTGTAGCTCCAGGACATTTTCAAGCCATTTACACTAGTTATTATGTATG 1829  
QY 55 -----Ser-----His-----His 57  
Db 1828 AGGTCTAGAAAACATGATTGAGCTGCATCCCAAGCACACGAGTGCAGTTAAGTTAGCAC 1769  
QY 58 Trp-----IleGlnGluSerLeuLeu-----Cys 65  
Db 1768 TGGAGTTGATTTTCTGTCTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1712  
QY 66 ---Pro-Pro-----Ser-----ProLys-----Glu---Val----- 72  
Db 1711 CATCCATCCCTCTTTTCTGTGTACAGGTCTGATATCCGAAGGCTGGGACGCTATGCAAGG 1652  
QY 73 -Thr-----Cys---ArgGlu-----Met---Leu----- 78  
Db 1651 GACCCAAAATTTGTGTGTCAGCAACAAACAGGAAATCGCTGATGCTGTGACCTCACCATC 1592

PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-541565/60.  
XX

```

QY 79 -Thr-----Gly-----Gly---CysLeu-----Pro-- 84
    |||
Db 1591 GACGAGATGGAGAGTGCAGCCAGCCCTGCTTAATGGGAACGTGCGTCCCGGAGCCAAC 1532
QY 85 -----TrpAla-----Thr---Arg-----SerHis----- 90
    |||||
Db 1531 GGGGATGTGGGCCCCCTCTCACACCGGCAGGACTATGAGCTACAGGACTTTGGTCCTGGC 1472
QY 91 -----Leu---Gly-----Arg-----Arg---Lys----- 95
    |||
Db 1471 TACAGCGACGAAGAGCCAGCCCTGGGAGGGATGAGGAGGACCTGGCGGATGAATGATA 1412
QY 96 -----CysSer 97
    |||||
Db 1411 TGCATCACCACCTTGTAGC 1393

```

Search completed: May 11, 2003, 18:48:35  
Job time : 161 secs



Tue May 13 12:12:50 2003

Sequence 26, Appli  
Sequence 3, Appli  
Sequence 24, Appli  
Sequence 1, Appli  
Sequence 24, Appli  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 58, Appli  
Sequence 58, Appli  
Sequence 58, Appli  
Sequence 95, Appli  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 28, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 86, Appli  
Sequence 167, Appli  
Sequence 9, Appli  
Sequence 9, Appli  
Sequence 2, Appli  
Sequence 302, Appli  
Sequence 1, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 7, Appli

US-09-018-635-26  
US-09-467-642-3  
US-09-041-886-24  
PCT-US94-05277-1  
US-09-724-864-24  
US-09-249-697A-2  
US-09-363-316B-2  
US-08-631-607-1  
US-09-098-358B-1  
US-09-232-200-58  
US-09-232-197-58  
US-09-232-201-58  
US-09-221-017B-995  
US-08-106-761-3  
US-09-436-983-1  
PCT-US94-01712-28  
US-09-013-895A-1  
US-09-565-918-1  
US-09-448-868-1  
US-09-058-489-86  
US-09-221-017B-167  
US-08-198-446B-9  
US-09-228-986-2  
US-09-149-476-302  
US-08-824-996-1  
US-09-999-811-3  
US-09-042-105-3  
US-08-999-811-1  
US-09-042-105-1  
PCT-US96-09001-1  
US-08-795-430-7

2907 4  
2907 4  
4608 4  
4608 5  
2395 4  
1611 4  
1611 4  
2569 1  
2569 4  
3098 4  
3098 4  
3098 4  
4400 4  
1969 1  
2017 4  
2274 1  
2274 5  
2274 5  
2152 4  
2152 4  
2152 4  
2983 3  
3247 4  
1647 1  
1647 2  
2568 4  
1454 4  
1525 2  
1526 2  
1526 2  
1674 3  
1674 3  
1674 5  
1997 3

52.5  
52.5  
52.5  
52.5  
52.3  
52.0  
52.0  
52.0  
52.0  
52.0  
52.0  
52.0  
51.7  
51.6  
51.5  
51.5  
51.4  
51.4  
51.4  
51.4  
51.4  
51.2  
51.2  
51.2  
51.1  
51.1  
51.1  
51.1  
51.1  
51.1  
51.1  
51.1

279.5  
279.5  
279.5  
279.5  
278.5  
276.5  
276.5  
276.5  
276.5  
276.5  
276.5  
276.5  
275  
274.5  
274  
274  
273.5  
273.5  
273.5  
273.5  
273.5  
272.5  
272.5  
272.5  
272  
272  
272  
272  
272  
272  
272  
272

C 13  
C 14  
C 15  
C 16  
C 17  
C 18  
C 19  
C 20  
C 21  
C 22  
C 23  
C 24  
C 25  
C 26  
C 27  
C 28  
C 29  
C 30  
C 31  
C 32  
C 33  
C 34  
C 35  
C 36  
C 37  
C 38  
C 39  
C 40  
C 41  
C 42  
C 43  
C 44  
C 45

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 18:46:17 ; Search time 42 seconds  
(without alignments)  
708.277 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHASLGDSLTSLQ.....LTGGCLPWATRSHLGRKCS 97

Scoring table:  
BLOSUM62  
Xgapop 0.0 , Xgapext 0.5  
Ygapop 0.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues  
Total number of hits satisfying chosen parameters: 875442

Minimum DB seq length: 0  
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/uspto\_spool/US09854133/runat\_05052003\_174426\_1235/app\_query.fasta\_1.263  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=ptc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=5000  
-USER=US09854133@cgn\_1\_1\_32@runat\_05052003\_174426\_1235 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=0 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=0 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	283	53.2	2533	4	US-08-861-745B-2
2	283	53.2	3520	1	US-08-424-788-1
3	283	53.2	3520	1	US-08-110-683-3
4	283	53.2	3520	2	US-08-477-166-3
5	283	53.2	3520	2	US-08-472-097-3
6	283	53.2	3520	4	US-09-439-672-3
7	283	53.2	3520	5	PCT-US93-11638-3
8	282.5	53.1	3025	1	US-08-444-734A-1
9	282	53.0	1001	4	US-09-641-638-319
10	281	52.8	2693	1	US-07-925-695-4
C 11	279.5	52.5	2907	2	US-09-018-628-17
C 12	279.5	52.5	2907	3	US-09-273-378-17

ALIGNMENTS

RESULT 1  
US-08-861-745B-2/c  
; Sequence 2, Application US/08861745B  
; Patent No. 6165733  
; GENERAL INFORMATION:  
; APPLICANT: Cen, Hui  
; APPLICANT: Williams, Lewis  
; TITLE OF INVENTION: Gamma II Adaptin  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/861,745B  
; FILING DATE: 22-MAY-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A 32141  
; REGISTRATION NUMBER: 02441.05336  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; TELEX:

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2533 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-861-745B-2

Alignment Scores:  
Pred. No.: 1.42e+04  
Score: 283.00  
Percent Similarity: 25.33  
Best Local Similarity: 32.05%  
Query Match: 24.36%  
DB: 53.20%  
Indels: 146  
Gaps: 51

US-09-854-133-586 (1-97) x US-08-861-745B-2 (1-2533)

```
QY 1 Glu---ValGluVal-----Ser---Arg-----Asp-----7
Db 2137 CAGTGACTAAGACAAAGCAGGGGTCTCCAAAGGCGCATGAAGAAGATCCAGCTGTA 2078
QY 8 -----HisAla---SerLeu-----Gly---Asp-----Ser-----14
Db 2077 GGCCCTCAGGCTCAACACTCTGACACTGGGATGGGAGCTGGGGTGGCGGTGTACAGG 2018
QY 15 GluThrLeuSerGln---ThrGluLeuArg-----LysLys---GluArgLys 28
Db 2017 GAA-----GGTCAAGGAGATGAATTAAGGCTTCCCTGGGAAGGAGTCTGAGGAGGAA 1964
QY 29 -----LysLysArgGluArgLysPheGlnAla---AsnCys-----Gly 40
Db 1963 GATGCTGGGCATGCCAGAGAGAG-----GCTCTGAAGTGTACCCAGGA 1916
QY 41 -----IleAspPheIle---IlePheTrp-----IlePheTrp-----Ile---51
Db 1915 GATCCAGTAGAT---CTAAGAGTTTGGTGGCTGGGTTCTGTGGGSCAGGGGTGTCA 1859
QY 52 Leu-----Leu---Phe-----Ser-----His 56
Db 1858 CTTCCAGGGCTGGGCTTCCGTTTGGTCTTCTCCTCCCTCCCTTTAAGCTGGGGTGCAC 1799
QY 57 HisTrpIleGlnGlu---Ser---Leu-----Leu-----64
Db 1798 CAGCTCTACAGAGGACATCTTTTCTAGGATGGGCGCTCATGTGGTCTCTCTCTGGA 1739
QY 65 -----Cys-----Pro---Pro-----SerPro-----Lys 70
Db 1738 AGAGTGTGTATTAATCCACAGCCCGCTGTTCGAGCTCTAAGTCCACACAGCTCCCGTAGA 1679
QY 71 -----GluValThrCysArgGluMetLeuThrGly---GlyCys-----82
Db 1678 TGGACACCACTGGCGAATACGATTGT-----TGTCTCCCGAGTGGGTGCTCAGCT 1625
QY 83 -----Leu-----Pro---Trp---Ala-----Thr-----87
Db 1624 TCATGAGGCTGTGTATGGCGTAGCCCGAGTGGCTGGCAGGACATATGGGACTGCACCA 1565
QY 88 -----Arg-----Ser-----His 90
Db 1564 CTTTTCAGCAGTGTGTACAGCTCTCTTCTTCCACCTGAAAGGCTCCGCTCTCTCTCAC 1505
QY 91 -----Leu-----Gly---Arg---ArgLysCysSer 97
Db 1504 AGTTACCTTCCAGCAGGAGGTCCCCATACCTCGCCATGCACC 1463
```

## RESULT 2

US-08-424-788-1  
Sequence 1, Application US/08424788  
Patent No. 5716804  
GENERAL INFORMATION:  
APPLICANT: Moore, Kevin W.  
APPLICANT: Wei, Sherry

APPLICANT: HO, Alice Suk-Yue  
TITLE OF INVENTION: MAMMALIAN INTERLEUKIN-10 (IL-10)  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DNAX Research Institute  
STREET: 901 California Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/424,788  
FILING DATE: 19-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0501  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-852-9196  
TELEFAX: 415-496-1200  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3520 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 80..1807  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 128..1807  
US-08-424-788-1

Alignment Scores:  
Pred. No.: 4.87e+04  
Score: 283.00  
Percent Similarity: 25.66%  
Best Local Similarity: 20.72%  
Query Match: 53.20%  
DB: 1  
Length: 3520  
Matches: 63  
Conservative: 15  
Mismatch: 10  
Indels: 216  
Gaps: 62

```
US-09-854-133-586 (1-97) x US-08-424-788-1 (1-3520)
QY 1 Glu-----ValGluValSerArg-----Asp-----HisAla 9
Db 2001 GAGTGAGGCTGTGTAGATACCAGCAGAGCTGAGCAGGATTGACAGACCTCTCATGCC 2060
QY 10 Ser-----Leu-----Gly-----12
Db 2061 TCAGGGCTGGCTCTACACTGGAAGGACCTGTGTTGGTGTAACTCAGGGCTTCTG 2120
QY 13 -----Asp-----SerGlu---Thr-----LeuSer---GlnThr---Glu 21
Db 2121 ATGTGGTAAGACTGTAGTCTGAAGTCAGCTGAGCTGGATGTCTGGGAGGTGTGGAG 2180
QY 22 ---Leu-----Arg---Lys-----LysGluArg---Lys-----29
Db 2181 TGGCTAGCTGTCTACAGGATAAAGGAGGCTCAAG---AGATAGAGGCGCAGCATGA 2237
QY 30 ---Lys-----Arg---Glu-----Arg-----33
Db 2238 GCCAGGTTAATTTTGTCTGTAGATGGTCCCGCAGGAGTGGTGTGGCTG 2297
QY 34 ---Lys-----Phe-----Gln-----AlaAsn-----Cys---39
```

us-09-854-133-586.rni

Tue May 13 12:12:50 2003

CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/011,066  
FILING DATE: 29-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0335K1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-852-9196  
TELEFAX: 415-496-1200  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3520 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 80..1807  
US-08-110-683-3

Alignment Scores:  
Pred. No.: 4.87e+04  
Score: 283.00  
Percent Similarity: 25.66%  
Best Local Similarity: 20.72%  
Query Match: 53.20%  
DB: 1  
Length: 3520  
Matches: 63  
Conservative: 15  
Mismatches: 10  
Indels: 216  
Gaps: 62

US-09-854-133-586 (1-97) x US-08-110-683-3 (1-3520)

QY 1 Glu-----ValGluValSerArg-----Asp-----HisAla 9  
Db 2001 GAGTGAGGGTCTGTAGATACACAGAGAGCTGAGCAGGATTGACAGAGACCTCCTCATGCC 2060  
QY 10 Ser-----Leu-----Gly----- 12  
Db 2061 TCAGGGCTGGCTCCTACACTGGAGGACCTGTGTTGGGTGTAACCTCAGGGCTTTCTG 2120  
QY 13 -----Asp-----SerGlu-----Thr-----LeuSer-----GlnThr 21  
Db 2121 ATGTGGTAAGACTGTAGTCTGAAGTCAAGCTGAGCCTGAGCTGGATGTCTGCGGAGGTGTGGAG 2180  
QY 22 -----Leu-----Arg-----Lys-----LysGluArg-----Lys----- 29  
Db 2181 TGCTAGCCTGCTACAGGATAAGGGAAGGCTCAAG-----AGATAGAGGGCAGAGCATGA 2237  
QY 30 -----Lys-----Arg-----Glu-----Arg----- 33  
Db 2238 GCCAGGTTTAATTTTGTCTGTAGAGATGGTCCCGCAGCCAGGATGGTTACTTGTGGCTG 2297  
QY 34 -----Lys-----Phe-----Gln-----AlaAsn-----Cys----- 39  
Db 2298 GGAGATCTTGGGTATACACCCCTGAATGATCAGCCAGTCAATTCAGAGCTGTGTGGC 2357  
QY 40 -----GlyIle----- 41  
Db 2358 AAAAGGACTGAGACCCAGAAATTTCTGTTCTCTCTGTGAGGTGTCTCTGTACCCATCTG 2417  
QY 42 -----Asp-----PheIleIlePhe-----Trp-----Ile-----Phe----- 50  
Db 2418 CAGACAGACATCTTC-----ATCTTTTACTATGGCTGTGTCTCCCTGAATTACCCAGCAGTGG 2474  
QY 51 -----IleLeuLeu----- 54  
Db 2475 CCAAGCCATTACTCCCTGCTGCTCACTGTGTGACGTGACAGACCCAGACGCTGTCTG 2534  
QY 55 Ser-----HisHis----- 60  
Db 2535 TCTGTGTAGTACACTACCTTTAGGTGGCTTTGGGCTTGAGCACTGGCCCGCAGGCTTAG 2594  
QY 61 Glu-----LeuLeu-----Cys----- 65  
Db 2595 GACTTATGTCTGCTTTGCTGCTAATCTCTAAGTCCCTCATGCTGAATGTTAACCAAGG----- 67  
QY 66 -----ProPro-----Ser-----Pro-----Lys 70  
Db 2655 CTGACACTCCGTTTCAGCTGTGTGACCTCCGACAGAGCTTCTCAGGGGACTAAAA 2714  
QY 71 -----GluVal-----Thr-----Cys-----ArgGluMetLeuThr 79  
Db 2715 TAATGACTAGTCAATTCAGAGAGTCCCTCATGCTGAATGTTAACCAAGG----- 87  
QY 80 GlyGlyCysLeuProTrp-----AlaThr----- 87  
Db 2763 -----TGC-----CCCTGGGGTGATAGTTTAGTCTGCAACCTCTGGGTGGAAGGAAGT 2813  
QY 88 -----Arg-----Ser-----HisLeu-----Gly-----ArgArg 94  
Db 2814 GGACTACGGAAGCCATCTGTCTCCCTGGGAGCTTCCACCTCATGCCAGTGTTCAGAGA 2873  
QY 95 LysCys-----Ser 97  
Db 2874 TCTTGTGGGAGC 2885

RESULT 3

US-08-110-683-3  
Sequence 3, Application US/08110683  
Patent No. 5789192

GENERAL INFORMATION:

APPLICANT: Moore, Kevin W.  
APPLICANT: Liu, Ying  
APPLICANT: Ho, Alice Suk-Yue  
APPLICANT: Hsu, Di-Hwei  
APPLICANT: Bazan, J. Fernando  
APPLICANT: Tan, Jimmy C.  
APPLICANT: Chou, Chuan-Chu  
TITLE OF INVENTION: Mammalian Receptors for Interleukin-10  
TITLE OF INVENTION: (IL-10)  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESS: DNAX Research Institute  
STREET: 901 California Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/110,683  
FILING DATE: 23-AUG-1993

```
; MOLECULE TYPE:   CDNA
; FEATURE:
; NAME/KEY:   CDS
; LOCATION:   80..1807
US-08-477-166-3

Alignment Scores:
Pred. No.:      4.87e+04
Score:          283.00
Percent Similarity: 25.66%
Best Local Similarity: 20.72%
Query Match:     53.20%
DB:              2
Gaps:            62

US-09-854-133-586 (1-97) x US-08-477-166-3 (1-3520)

QY    1 Glu-----ValGluValSerArg-----Asp-----HisAla 9
|||   |||::: |||||| |
Db    2001 GAGTCAGGGTCTGTAGATACCACGACAGCTGAGCAGGATTGCACAGAGACCTCCTCATGCC 2060
|||   |||::: |||||| |

QY    10 Ser-----Leu-----Gly----- 12
|||   |||   |||   |||
Db    2061 TCAGGGCTGGCTCCTACACTGGAAGGACCTGTGTTGGGTGAACCTCAGGGCTTTCTGG 2120
|||   |||   |||   |||

QY    13 -----Asp-----SerGlu---Thr-----LeuSer---GlnThr---Glu 21
|||   |||   |||   |||::: |||   |||   ::|
Db    2121 ATGTGGTAAGACTGTAGGTCTGAAGTCAGCTGAGCCTGGATGCTCGGAGGTGTGGAG 2180
|||   |||   |||   |||   |||   |||   |||   |||

QY    22 ---Leu-----Arg---Lys-----LysGluArg---Lys---Lys--- 29
|||   |||   |||   |||   |||   |||   |||   |||   ::|
Db    2181 TGGCTAGCCTGCTACAGGATAAAGGAAGGCTCAAG---AGATAGAAGGCAGAGCATGA 2237
|||   |||   |||   |||   |||   |||   |||   |||

QY    30 ---Lys-----Lys-----Arg---Glu---Arg----- 33
:::   |||   |||   |||   |||
Db    2238 GCCAGGTTTAATTTTGCTCTGTAGAGATGGTCCCGCAGCAGGATGGTTACTTGTGGCTG 2297
|||   |||   |||   |||   |||

QY    34 ---Lys-----Phe-----Gln---AlaAsn---Cys--- 39
|||   |||   |||   |||   |||
Db    2298 GGAGATCTTGGGGTATACACCACCCTGAATGATCAGCCAGTCAATTACAGAGCTGTGTGGC 2357
|||   |||   |||   |||   |||

QY    40 -----GlyIle----- 41
|||   |||   |||

Db    2358 AAAAGGGACTGAGACCCAGAAATTTCTGTTCTCTTGTGAGGTGCTCTGCTACCCATCTG 2417
|||   |||   |||   |||

QY    42 -----Asp---PheIleIlePhe---Trp-----Ile---Phe---Trp 50
|||   |||   |||   |||   |||   |||   |||   |||   ::|
Db    2418 CAGACAGACATCTTC---ATCTTTTACTATGGCTGTGTCCCGTGAATTACCCAGCAGTGG 2474
|||   |||   |||   |||   |||

QY    51 -----IleLeuLeu-----Cys----- 65
|||   |||   |||   |||
Db    2475 CCAAGCCATTACTCCCTGCTGCTCACTGTTGTGAGCTCAGACCAGACGAGCTGTCTG 2534
|||   |||   |||   |||

QY    55 Ser-----HisHis-----Trp-----Ile---Gln----- 60
|||   |||   |||   |||   |||   |||
Db    2535 TCTGTGTTAGTACTACCTACCTTTTAGGTGGCCCTTGGGCTTGAGCACTGGCCCAGGCTTAG 2594
|||   |||   |||   |||

QY    61 Glu---Ser-----LeuLeu-----Cys----- 65
|||   |||   |||   |||
Db    2595 GACTTATGTCTGCTTTTGTGCTGCTAATCTCTAACTGCAGACCCAGAGAACAGGGTGTCTGG 2654
|||   |||   |||   |||

QY    66 -----ProPro---Ser-----Pro-----Lys 70
|||   |||   |||   |||
Db    2655 CTGACACCTCCGTTTCAGCTGTGTGACCTCCGACCCAGCAGCTTCCTCAGGGGACTAAAA 2714
|||   |||   |||   |||

QY    71 -----GluVal-----Thr---Cys-----ArgGluMetLeuThr 79
|||   |||   |||   |||
Db    2715 TAATGACTAGGTCAATCAGAAGTCCCTCATGCTGAATGTTAACCACAGG----- 2762
|||   |||   |||   |||

QY    80 GlyGlyCysLeuProTrp-----AlaThr----- 87
|||   |||   |||   |||
Db    2763 -----TGC---CCCTGGGGTGATAGTTTAGGTCTCTGCAACCTCTGGTGGTGAAGGAAGT 2813
|||   |||   |||   |||

QY    88 -----Arg-----Ser---HisLeu---Gly-----ArgArg 94
|||   |||   |||   |||
```



Db 2814 GGACTACGGAAGCCATCTGTCCCCCTGGGGAGCTTCCACCTCATGCGAGTGTTCAGAGA 2873

QY 95 LysCys---Ser 97  
Db 2874 TCTTGTGGGAGC 2885

RESULT 5  
US-08-472-097-3  
; Sequence 3, Application US/08472097  
; Patent No. 5985828  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Kevin W.  
; APPLICANT: Liu, Ying  
; APPLICANT: HO, Alice Suk-Yue  
; APPLICANT: Hsu, Di-Hwei  
; APPLICANT: Bazan, J. Fernando  
; APPLICANT: Tan, Jimmy C.  
; APPLICANT: Chou, Chuan-Chu  
; TITLE OF INVENTION: Mammalian Receptors for Interleukin-10  
; TITLE OF INVENTION: (IL-10)  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472,097  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,683  
; FILING DATE: 23-AUG-1993  
; APPLICATION NUMBER: US 08/011,066  
; FILING DATE: 29-JAN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; REFERENCE/DOCKET NUMBER: DX0335K1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-852-9196  
; TELEFAX: 415-496-1200  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3520 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 80..1807  
; US-08-472-097-3

Alignment Scores: 4.87e+04 3520  
Pred. No.: 283.00  
Score: 25.66%  
Percent Similarity: 20.72%  
Best Local Similarity: 53.20%  
Query Match: 2  
DB: 62

US-09-854-133-586 (1-97) x US-08-472-097-3 (1-3520)

QY 1 Glu-----ValGluValSerArg-----Asp-----HisAla 9

Db 2001 GAGTGAGGGTCTGTAGATACACAGCAGAGCTGAGCAGGATTGACAGAGACCTCCTCATGCC 2050  
QY 10 Ser-----Leu-----Gly----- 12  
Db 2061 TCAGGGCTGCTCTACACTGGAAGGACCTGTGTTGGGTGTAACCTCAGGGCTTCTCTGG 2120  
QY 13 -----Asp-----SerGlu---Thr-----LeuSer---GlnThr---Glu 21  
Db 2121 ATGTGGTAAGACTGTAGGTCTGAAGTCAGCTGAGCCTGGATGTCTCGGAGGTTGGAG 2180  
QY 22 ---Leu-----Arg---Lys-----LysGluArg---Lys-----Lys 29  
Db 2181 TGGCTAGCCTGCTACAGGATAAAGGGAAGGCTCAAG---AGATAGAAGGGCAGACCATGA 2237  
QY 30 ---Lys-----Arg---Glu-----Arg----- 33  
Db 2238 GCCAGGTTTAATTTTGTCTGTAGAGATGGTCCCCAGCAGGATGGGTACTTGTGGCTG 2297  
QY 34 ---Lys-----Phe-----Gln---AlaAsn---Cys--- 39  
Db 2298 GGAGATCTTGGGTATATACACACCCTGAATGATCAGCCAGTCAATTCAGAGCTGTGTGGC 2357  
QY 40 -----GlyIle----- 41  
Db 2358 AAAAGGGACTGAGACCCAGAAATTTCTGTCTCTTGTGAGGTGTCTCTGTACCCATCTG 2417  
QY 42 -----Asp---PheIleIlePhe---Trp---Ile---Phe---Trp 50  
Db 2418 CAGACAGACATCTTC---ATCTTTTACTATGGCTGTGTCCCTGTAATTAACAGCAGTGG 2474  
QY 51 -----IleLeuLeu-----Phe 54  
Db 2475 CCAAGCCATTACTCCCTGCTGCTCACTGTGTGACGTCAGACCCAGACAGCAGCTGTCTG 2534  
QY 55 Ser-----HisHis-----Trp-----Ile---Gln--- 60  
Db 2535 TCTGTGTAGTACACTACCTTTAGGTGGCTTTGGGCTTGAGCACTGGCCAGGCTTAG 2594  
QY 61 Glu---Ser-----LeuLeu-----Cys----- 65  
Db 2595 GACTTATGTCTGCTTTTGTCTGCTAATCTCTAATCTGACACCCAGACAGAGAGGCTGGG 2654  
QY 66 -----ProPro---Ser-----Pro-----Lys 70  
Db 2655 CTGACACCTCCGTTGTCAGCTGTGTGACCTCCGACCCAGCAGCTTCTCTCAGGGGACTAAA 2714  
QY 71 -----GluVal-----Thr---Cys-----ArgGluMetLeuThr 79  
Db 2715 TAATGACTAGGTCAATTCAGAAGTCCCTCATGCTGAATGTTAACAAGG----- 2762  
QY 80 GlyGlyCysLeuProTrp-----AlaThr----- 87  
Db 2763 -----TGC---CCCTGGGTGATAGTTTAGTCTGCAACCTCTGGGTTGGAAGGAAGT 2813  
QY 88 -----Arg-----Ser---HisLeu---Gly-----ArgArg 94  
Db 2814 GGACTACGGAAGCCATCTGTCCCTGGGGAGCTTCCACCTCATGCCAGTGTTCAGAGA 2873  
QY 95 LysCys---Ser 97  
Db 2874 TCTTGTGGGAGC 2885

RESULT 6  
US-09-439-672-3  
; Sequence 3, Application US/09439672  
; Patent No. 6423500  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Kevin W.  
; APPLICANT: Liu, Ying  
; APPLICANT: HO, Alice Suk-Yue  
; APPLICANT: Hsu, Di-Hwei  
; APPLICANT: Bazan, J. Fernando

APPLICANT: Tan, Jimmy C.  
APPLICANT: Chou, Chuan-Chu  
TITLE OF INVENTION: Mammalian Receptors for Interleukin-10 (IL-10)  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DNAX Research Institute  
STREET: 901 California Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/439,672  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/110,683  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0335K1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-852-9196  
TELEFAX: 415-496-1200  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3520 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 80..1807  
US-09-439-672-3

Alignment Scores:

Pred. No.: 4.87e+04  
Score: 283.00  
Percent Similarity: 25.66%  
Best Local Similarity: 20.72%  
Query Match: 53.20%  
DB: 4  
Length: 3520  
Matches: 63  
Conservative: 15  
Mismatch: 10  
Indels: 216  
Gaps: 62

US-09-854-133-586 (1-97) x US-09-439-672-3 (1-3520)

QY 1 Glu-----ValGluValSerArg-----Asp-----Hisala 9  
Db 2001 GAGTGAGGGTCTGTAGATACACAGAGCTGAGCAGGATTGACAGAGACCTCCTCATGCC 2060  
QY 10 Ser-----Leu-----Gly-----12  
Db 2061 TCAGGGCTGGCTCCTACACTGGAAGGACCTGTGTTGGGTGAACCTCAGGGCTTCTG 2120  
QY 13 -----Asp-----SerGlu---Thr-----LeuSer---GlnThr---Glu 21  
Db 2121 ATGTGGTAAGACTGTAGGTCTGAAGTCAGCTGAGCCTGGATGTCTGCGGAGGTGTGGAG 2180  
QY 22 ---Leu-----Arg---Lys-----LysGluArg---Lys---Lys-----29  
Db 2181 TGGCTAGCCTGCTACAGGATAAAGGGAAGGCTCAAG---AGATAGAAGGCGAGCATGA 2237  
QY 30 ---Lys-----Arg---Glu-----Arg-----33  
Db 2238 GCCAGGTTAAATTTGTCTGTAGAGATGGTCCCCAGGATGGGTTACTTGTGGCTG 2297

QY 34 ---Lys-----Phe-----Gln---AlaAsn-----Cys---39  
Db 2298 GGAGATCTTGGGTATACACACCCTGAATGATCAGCCAGTCAATTACAGAGCTGTGGC 2357  
QY 40 -----Glyile-----41  
Db 2358 AAAAGGAGTACAGACCCAGAAATTTCTGTTCTCTTGTGAGGTGTCTCTGCTACCCATCTG 2417  
QY 42 -----Asp---PheIleIlePhe-----Trp-----Ile---Phe---Trp 50  
Db 2418 CAGACAGACATCTTC---ATCTTTTACTATGGCTGTGCTCCCTGAATTACACAGCAGTGG 2474  
QY 51 -----IleLeuLeu-----Phe 54  
Db 2475 CCAAGCCATTACTCCCTGCTGCTCACTGTTGTGACGTGACACACAGACAGACGCTGTCTG 2534  
QY 55 Ser-----HisHis-----Trp-----Ile---Gln-----60  
Db 2535 TCTGTGTTAGTACACTACCCCTTAGGTGGCCTTTGGGCTTGAGCAGTGGCCAGGCTTAG 2594  
QY 61 Glu---Ser-----LeuLeu-----Cys-----65  
Db 2595 GACTTATGCTGCTTGTGCTGCTAATCTCTAATCTGACACCCAGAGAACAGGGTGTCTGG 2654  
QY 66 -----ProPro---Ser-----Pro-----Lys 70  
Db 2655 CTGACACCTCCGTTTCAGCTGTGTGACCTCCGACACAGAGTTCCTCAGGGGACTAAAA 2714  
QY 71 -----GluVal-----Thr---Cys-----ArgGluMetLeuThr 79  
Db 2715 TAATGACTAGGTCAATTCAGAGAGTCCCTCATGCTGAATGTTAACCAGG-----2762  
QY 80 GlyGlyCysLeuProTrp-----AlaThr-----87  
Db 2763 -----TGC---CCCTGGGGTGATAGTTAGGTCTGCAACCTCTGGGTGGAAGAGT 2813  
QY 88 -----Arg-----Ser---HisLeu---Gly-----ArgArg 94  
Db 2814 GGACTACGGAAGCCATCTGTCCCTCCCTGGGAGCTTCCACCTCATGCCAGTGTTCAGAGA 2873  
QY 95 LysCys---Ser 97  
Db 2874 TCTTGTGGGAGC 2885  
RESULT 7  
PCT-US93-11638-3  
; Sequence 3, Application PC/TUS9311638  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Mammalian Receptors For Interleukin-10 (IL-10)  
; NUMBER OF SEQUENCES: 12  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Macintosh 6.0.8  
; SOFTWARE: Microsoft Word 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/11638  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,683  
; FILING DATE: 23-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/011,066  
; FILING DATE: 29-JAN-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/989,792  
; FILING DATE: 10-DEC-1992  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3520 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single

us-09-854-133-586.rni

Tue May 13 12:12:50 2003

TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
PCT-US93-11638-3

Alignment Scores:  
Pred. No.: 4.87e+04 Length: 3520  
Score: 283.00 Matches: 63  
Percent Similarity: 25.66% Conservativity: 15  
Best Local Similarity: 20.72% Mismatches: 10  
Query Match: 53.20% Indels: 216  
DB: 5 Gaps: 62

US-09-854-133-586 (1-97) x PCT-US93-11638-3 (1-3520)

QY	1	Glu	-----ValGluValSerArg-----HisAla 9
Db	2001	GAGTGAGGGTCTGTAGATACACGAGAGCTGAGCAGGATTGACAGAGACCTCCTCATGCC 2060	
QY	10	Ser	-----Leu-----Gly-----12
Db	2061	TCAGGGCTGGCTCTACACTGGAAGACCTGTGTTGGGTGTAACTCAGGGCTTCTGTG 2120	
QY	13	-----Asp-----SerGlu-----Thr-----LeuSer-----GlnThr-----Glu 21	
Db	2121	ATGTGGTAAGACTGTAGGTCTGAAGTCAGCTGAGCCTGGATGCTGCGGAGGTGTGGAG 2180	
QY	22	-----Leu-----Arg-----Lys-----LysGluArg-----Lys-----Lys-----29	
Db	2181	TGGCTAGCCTGCTACAGGATAAAGGGAAGGCTCAAG-----AGATAGAGGGCAGAGCATGA 2237	
QY	30	-----Lys-----Arg-----Glu-----Arg-----33	
Db	2238	GCAGGTTTAATTTTGTCTGTAGAGATGTTCCCGAGGATGGTTACTTGTGGCTG 2297	
QY	34	-----Lys-----Phe-----Gln-----AlaAsn-----Cys-----39	
Db	2298	GGAGATCTTGGGTATACACACCCCTGAATGATCAGCCAGTCAATTCAGAGCTGTGTGGC 2357	
QY	40	-----Glylle-----41	
Db	2358	AAAAGGGACTGAGACCCAGAAATTTCTGTCTCTTGTGAGGTGTCTGTACCCATCTG 2417	
QY	42	-----Asp-----PheIleIlePhe-----Trp-----Ile-----Phe-----Trp 50	
Db	2418	CAGACAGACATCTC-----ATCTTTTACTATGCTGTGCTCCCTGAATTACCAGCAGTGG 2474	
QY	51	-----IleLeuLeu-----Phe 54	
Db	2475	CCAAGCCATTACTCCCTGCTGCTCACTGTTGTGAGCTCAGACCAGACGACGCTGTCTG 2534	
QY	55	Ser-----HisHis-----Trp-----Ile-----Gln-----60	
Db	2535	TCTGTGTTAGTACACTACCTTTTAGGTGGCTTTGGGCTTGAGCACTGGCCCCAGGCTTAG 2594	
QY	61	Glu-----Ser-----LeuLeu-----Cys-----65	
Db	2595	GACTTATGCTGCTTTTGTCTGCTAATCTTAAGTGCAGACCCAGAGAACAGGGTGTGGG 2654	
QY	66	-----ProPro-----Ser-----Pro-----Lys 70	
Db	2655	CTGACACCTCCGTGTTTCAGCTGTGTGACCTCCGACAGCAGCTTCTCAGGGGACTAAA 2714	
QY	71	-----GluVal-----Thr-----Cys-----ArgGluMetLeuThr 79	
Db	2715	TAATGACTAGGTCAATTCAGAAGTCCCTCATGCTGAATGTTAACCAGG-----2762	
QY	80	GlyGlyCysLeuProTrp-----AlaThr-----87	
Db	2763	-----TGC-----CCCTGGGGTGATGTTAGTCTCTGCAACCTCTGGTGGGAAGGAGT 2813	
QY	88	-----Arg-----Ser-----HisLeu-----Gly-----ArgArg 94	
Db	2814	GGACTACGGAAGCCATCTGTCCCCCTGGGGAGGTTCCACCTCATGCCAGTGTTCAGAGA 2873	

QY 95 LysCys---Ser 97  
Db 2874 TCTGTGGGAGC 2885

RESULT 8  
US-08-444-734A-1  
; Sequence 1, Application US/08444734A  
; Patent No. 5610282  
; GENERAL INFORMATION:  
; APPLICANT: Sibley, David R.  
; APPLICANT: Monsma, Lawrence C.  
; APPLICANT: McVittie, Loris D.  
; TITLE OF INVENTION: CDNA encoding the rat D1 dopamine  
; TITLE OF INVENTION: receptor linked to adenylyl cyclase activation and  
; TITLE OF INVENTION: expression of the receptor protein in plasmid-transfected  
; TITLE OF INVENTION: cell lines  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson and Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/444,734A  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/029,917  
; FILING DATE: 03-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/548,714  
; FILING DATE: 06-JUL-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E.  
; REGISTRATION NUMBER: 34,115  
; REFERENCE/DOCKET NUMBER: NIH065.001FW1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (714) 760-0404  
; TELEFAX: (714) 760-9502  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3025 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: PB73D1  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 293..1756  
; US-08-444-734A-1

Alignment Scores:  
Pred. No.: 2.88e+04 Length: 3025  
Score: 282.50 Matches: 57  
Percent Similarity: 29.73% Conservativity: 20  
Best Local Similarity: 22.01% Mismatches: 11  
Query Match: 53.10% Indels: 171  
DB: 1 Gaps: 62

US-09-854-133-586 (1-97) x US-08-444-734A-1 (1-3025)







US-09-854-133-586 (1-97) x US-09-018-628-17 (1-2907) Gaps:

QY	1	GluValGlu-----ValSerArgAspHisAla-----SerLeu-----Gly-----AspSer	14
Db	1798	GAGACGGAGGTCGCAGTCGACCGAGATCACGCCACTGCACCTCCAGCCTGGGTGACAGAGC	1739
QY	15	GluThrLeuSerGlnThrGluLeuArgLysLysGluArgLysLysLysArgGluArg---	33
Db	1738	GAGACTCTGTCTCAA-----AAAAA-----AAAAAGAAAAAGAAAGAGCA	1691
QY	33	-----	---
Db	1690	GACTATCAGGGGCTATTATTAGGAACCATGCTCCTGTGAATCTGTGGAAATGAAAGCCT	1631
QY	34	-----Lys-----PheGln-----Ala-----Asn-----Cys	39
Db	1630	GTITCAGTTCAATGCCAAGCTTTTCATGGTCCGCCAGCGATCCTTAATCATCATCAGCTGT	1571
QY	40	-----Gly-----	---
Db	1570	TCGGTTAACAAATGGGTAATTTTAGAATGGCAGCCAGTTTCCTTCCCCATATTTCTG	1511
QY	42	-----Asp-----Phe-----IleIlePhe---TrpIle	48
Db	1510	CACTCCAGCCTTGACCCCACTCGCTTCTTCTACAGTCCACTCTGCTTTTGTATAT	1451
QY	49	-----Phe---Trp-----Ile---Leu-----LeuPheSerHis	56
Db	1450	GGTTGTACTGTCTTCATCTGGTGTGCTGCCTGAACCTGAAACAGTTTCATCCTCTTC	1394
QY	57	-----	---
Db	1393	CCAAGTCTCCTTTCTTCAACCCCATTAGAGCTGTTCCACTTGCCTTTGGGTACTTTGGG	1334
QY	58	-----Trp-----Ile-----	---
Db	1333	ATTCTTCTCTCCAGGGAGGGGTTGGTTGAGAACGGTGGCTTGGATGGTGGCGCTGAAGC	1274



Best Local Similarity:	20.57%	Mismatches:	6
Query Match:	52.54%	Indels:	258
DB:	4	Gaps:	53

US-09-854-133-586 (1-97) x US-09-018-635-26 (1-2907)

QY	1	GluValGlu	-----ValSerArgAspHisAla	---SerLeu	-----Gly	---AspSer	14
Db	1798	GAGACGGAGGTCGCAGTGAGCCGAGATCACGCCACTGCACCTCCAGCCTGGGTGACAGAGC					1739
QY	15	GluThrLeuSerGlnThrGluLeuArgLysLysGluArgLysLysLysArgGluArg					33
Db	1738	GAGACTCTGTCTCAA	-----AAAAA	-----AAAAA	-----AAAAA	-----AAAAA	1691
QY	33	-----					33
Db	1690	GACTATCAGGGGCTATTATTAGGAACCATGCTCCTGTGAATCTGTGGAAATGAAAGCCT					1631
QY	34	-----Lys	---PheGln	-----Ala	-----Asn	-----Cys	39
Db	1630	GTTTCAGTTCATGCCAAGTCTTTTCATGGTCCGCCAGCGATCCTTAATCATCATCAGCTGT					1571
QY	40	-----	-----Gly	-----Ile	-----Ile	-----Ile	41
Db	1570	TCGGTTAACAAATGGGTAATTTTAGAAATGGCAGCCAGTTTCCTTCCCCATATTCTG					1511
QY	42	-----Asp	-----Phe	-----IleIlePhe	-----TrpIle	-----TrpIle	48
Db	1510	CACCTCCAGCCTTGACCCACTCGCTTCTTCTACAGTCCACTTCTGCTCTTTTGTATATT					1451
QY	49	-----Phe	-----Trp	-----Ile	-----Leu	-----LeuPheSerHis	56
Db	1450	GGTTGTACTGTCTCATCTGGTGTGCCTGAACTTGAAACAGTTCATCCTCTTC					1394
QY	57	-----	-----His	-----His	-----His	-----His	57
Db	1393	CCAAGTCTCCTTTTCTTCAACCCCATTAGAGCTGTCCACTTGCCCTTTGGGTACTTTGGG					1334
QY	58	-----Trp	-----Ile	-----Ile	-----Ile	-----Ile	59
Db	1333	ATTCTTCTCCAGGGAGGGTTGGTTGAGAACGGTGGGCTTGGATGGTGGCGCTGAAGC					1274
QY	60	-----GlnGlu	-----Ser	-----Ser	-----Ser	-----Ser	62
Db	1273	GGCTCTCTGGGAGGAGTTGAGGCCCTGGCTGGGCTCAGTACTCTGGCTGTCTCTCCCTCAA					1214
QY	63	-----Leu	-----Leu	-----Cys	-----ProProSerPr	-----ProProSerPr	69
Db	1213	GACCAATCTGCTTATTGTATGTCATGCGCTTGTCTTGGGTGCAGTTCGAGCCACCTCACC					1154
QY	69	oLys	-----Glu	-----Glu	-----Glu	-----Glu	71
Db	1153	GTCAGCCGGGCTGAACCTTCGTTTTCATCTTTCTGGGTCTCTTGTTTTGAGGGCTGG					1094
QY	72	-----Val	-----Thr	-----Thr	-----Thr	-----Thr	73
Db	1093	TGATGCTGGAGAGCTTGAGTAGGAAGAACCATGATCCTTCTGGTCCAGTTTTCGAAAGGC					1034
QY	74	-----Cys	-----Arg	-----Glu	-----Glu	-----Glu	76
Db	1033	TGCCTCAGAATCCTGTGCACCAGACAGAGTCTTGAAGCTGCTTTCAGAGTCATCATTC					974
QY	77	Met	-----Leu	-----Leu	-----Leu	-----Leu	79
Db	973	AATGGTGGTTGGAGGATTCCGTAGCTGCCTTGCCTTCTGGTGGCTTTTCCACAGG					914
QY	80	-----Gly	-----Pro	-----Trp	-----Ala	-----Thr	88
Db	913	CCCTGGTGGCTGTTTATCTTCTTCCCTGTACTTGAGGCAGCGGACTCAGATTTCAA					854
QY	88	gSerHisLeu	-----Gly	-----Arg	-----Arg	-----Arg	95
Db	853	AGCCTTTTGGCCATCGTGAGGAGGTAGGGCTCGGCGTATCCAGGTGGCTCTCCAGGAA					794





```
Db 1458 CTGAAGGACCCAGGCTGTGTGTTATTCATGCTCGTTCCTGTGTGTACCTTCTCTGGA 1399
QY 67 -----Pro-----SerPro----- 69
Db 1398 GAAAAAGACCGTGAAAGTTTGAATGTTCCCTTTCGCTTCTGCAGGTGGCGCCAGCTGAG 1339
QY 70 -----Lys-----Glu---Val----- 72
Db 1338 ACGACAAATCGGCTGGAACCAAGACAGGACCACATCTCTGGGAGCCGAAGGAGGAC 1279
QY 73 Thr-----Cys-----Arg-----GluMetLeu---ThrGly---Gly--- 81
Db 1278 ACTGGAGCTGGGATTCAGGCTTAGGACAAATGAG---CTGTGCACTGGTCTGGGCATT 1222
QY 82 -----Cys-----Leu-----Pro--- 84
Db 1221 TCCAGCCTCATTTTCAGCCACACATTGATAAAGCCTTCATCTGACTTCACCAACCCCAAG 1162
QY 85 Trp-----Ala-----ThrArg-----Ser 89
Db 1161 TATCCGTAAGTTGCTTCCTCCCACTATCTGAAAATAATCACTAGGAATGACCACATCTCC 1102
QY 90 -----His---Leu-----Gly---Arg-----ArgLysCysSer 97
Db 1101 ATTCTTCATCCCAATTCACAGTGGGCACAGGCTTCCAGAGACTGTACA 1054
```

Search completed: May 11, 2003, 19:33:35  
Job time : 49 secs

Tue May 13 12:12:50 2003

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 19:06:36 ; Search time 90 Seconds  
(without alignments)  
1339.360 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHASLGDSITLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62  
Xgapop 0.0 , Xgapext 0.5  
Ygapop 0.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delopt 6.0 , Delext 7.0

Searched: 783854 seqs, 621352466 residues  
Total number of hits satisfying chosen parameters: 1556566

Minimum DB seq length: 0  
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV=xlh  
-MODEL=frame+p2n.model -USPTO=US09854133/runat\_05052003\_174428\_1338/app\_query.fasta\_1.263  
-Q=/cgn2\_1/USPTO.spool/US09854133/runat\_05052003\_174428\_1338/app\_query.fasta\_1.263  
-DB=PublishedApplications\_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosu62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=5000 -USER=US09854133.ecgn\_1.1.91.arunat\_05052003\_174428\_1338 -NCPU=6  
-ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=0 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=0 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	532	100.0	337	9 US-09-854-133-442	Sequence 442, App
2	532	100.0	337	10 US-09-738-973-442	Sequence 442, App
3	523	98.3	2239	9 US-09-854-133-440	Sequence 440, App
4	523	98.3	2239	10 US-09-738-973-440	Sequence 440, App

5	339.5	63.8	1861	9	US-10-163-866-30	Sequence 30, Appl
6	339.5	63.8	2482	9	US-10-163-866-29	Sequence 29, Appl
7	339	63.7	572	9	US-10-046-935-1307	Sequence 1307, Ap
8	339	63.7	572	9	US-09-878-178-1307	Sequence 1307, Ap
9	339	63.7	572	9	US-10-146-502-1307	Sequence 1307, Ap
10	298.5	56.1	1687	9	US-09-764-891-6421	Sequence 6421, Ap
11	298.5	56.1	1688	9	US-09-764-891-6422	Sequence 6422, Ap
12	296.5	55.7	2000	9	US-10-163-866-34	Sequence 34, Appl
13	292.5	55.0	3995	10	US-09-919-497-9	Sequence 9, Appl
14	287.5	54.0	2427	10	US-09-254-783A-2	Sequence 2, Appl
15	287.5	54.0	2427	12	US-10-152-058-2	Sequence 2, Appl
16	284.5	53.5	3243	9	US-10-092-154-1176	Sequence 1176, Ap
17	284.5	53.5	3243	10	US-09-764-847-1176	Sequence 1176, Ap
18	283.5	53.3	3162	10	US-09-764-877-2937	Sequence 2937, Ap
19	281.5	52.9	2203	9	US-10-224-562-1	Sequence 1, Appl
20	281.5	52.9	2203	10	US-09-801-861-1	Sequence 3, Appl
21	281.5	52.9	3308	9	US-10-147-026-3	Sequence 1754, Ap
22	281.5	52.9	4344	10	US-09-880-107-1754	Sequence 27, Appl
23	281	52.8	4168	12	US-10-042-417-27	Sequence 26, Appl
24	279.5	52.5	2907	10	US-09-912-962-26	Sequence 121, App
25	279.5	52.5	4870	12	US-10-044-090-121	Sequence 1138, Ap
26	279	52.4	3351	9	US-10-092-154-1138	Sequence 1138, Ap
27	279	52.4	3351	10	US-09-764-847-1138	Sequence 71, Appl
28	278.5	52.3	2281	9	US-10-071-766-71	Sequence 590, App
29	278.5	52.3	2395	9	US-10-152-661-590	Sequence 590, App
30	278.5	52.3	2395	9	US-09-866-050A-590	Sequence 1255, Ap
31	278	52.3	2495	9	US-10-091-504-1255	Sequence 1255, Ap
32	278	52.3	2495	10	US-09-764-869-1255	Sequence 9, Appl
33	278	52.3	2994	10	US-09-728-628-9	Sequence 397, App
34	277.5	52.2	2507	10	US-09-925-301-397	Sequence 3380, App
35	277.5	52.2	2808	10	US-09-880-107-3380	Sequence 179, App
36	277	52.1	2969	9	US-09-954-531-179	Sequence 2, Appl
37	276.5	52.0	1611	9	US-10-124-986-2	Sequence 2, Appl
38	276.5	52.0	1611	10	US-09-981-649A-2	Sequence 1, Appl
39	276.5	52.0	2569	9	US-10-224-951-1	Sequence 3199, Ap
40	276.5	52.0	2853	10	US-09-764-877-3199	Sequence 3196, Ap
41	276.5	52.0	2872	10	US-09-764-877-3196	Sequence 138, App
42	276.5	52.0	2903	9	US-09-984-827-138	Sequence 359, App
43	276.5	52.0	4152	10	US-09-925-300-359	Sequence 1443, Ap
44	276.5	52.0	4766	9	US-09-764-868-1443	Sequence 1568, Ap
45	276	51.9	2018	9	US-10-091-504-1568	

ALIGNMENTS

RESULT 1  
US-09-854-133-442  
; Sequence 442, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 442  
; LENGTH: 337  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-854-133-442

Alignment Scores: 1.77e-06 Length: 337  
Pred. No.: 532.00 Matches: 97  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00%

Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-09-854-133-586 (1-97) x US-09-854-133-442 (1-337)  
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
Db 5 GAGGTTGAAGTGACGACAGATCATGCCAGCTGGTGACAGTGAGACTGTCTCAACA 64  
QY 21 GluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 40  
Db 65 GAATTAAGGAAAAAGAAAGAAAGAAAGAGAGAGAGAGAAATCCAGGCCAATTGTGGC 124  
QY 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisTrpIleGln 60  
Db 125 ATAGATTTTATCATATTTTGGATTTTGGATTTTGGATTTTGGATTTTGGATTTTGG 184  
QY 61 GluSerLeuLeuCysProSerProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
Db 185 GAAAGCCTGTTGTGTCACCATCTCCAAAGGAGGTACCTGCAGGGAATGTTAAGGGA 244  
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
Db 245 GGCTGCCCTTCCTGGGCAACAAGGAGGACCTGGGCAAGGAAAGTGCAGC 295  
RESULT 2  
US-09-738-973-442  
; Sequence 442, Application US/09738973  
; Patent No. US20020110563A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Indirias, Carol Yoseph  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Elliot, Mark  
; APPLICANT: Mannion, Jane  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C9  
; CURRENT APPLICATION NUMBER: US/09/738,973  
; CURRENT FILING DATE: 2000-12-14  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 442  
; LENGTH: 337  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-738-973-442

Alignment Scores:  
Pred. No.: 1.77e-06 Length: 337  
Score: 532.00 Matches: 97  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 10 Gaps: 0  
US-09-854-133-586 (1-97) x US-09-738-973-442 (1-337)  
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
Db 5 GAGGTTGAAGTGACGACAGATCATGCCAGCTGGTGACAGTGAGACTGTCTCAACA 64  
QY 21 GluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 40  
Db 65 GAATTAAGGAAAAAGAAAGAAAGAAAGAGAGAGAGAGAAATCCAGGCCAATTGTGGC 124

QY 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisTrpIleGln 60  
Db 125 ATAGATTTTATCATATTTTGGATTTTGGATTTTGGATTTTGGATTTTGGATTTTGG 184  
QY 61 GluSerLeuLeuCysProSerProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
Db 185 GAAAGCCTGTTGTGTCACCATCTCCAAAGGAGGTACCTGCAGGGAATGTTAAGGGA 244  
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
Db 245 GGCTGCCCTTCCTGGGCAACAAGGAGGACCTGGGCAAGGAAAGTGCAGC 295  
RESULT 3  
US-09-854-133-440  
; Sequence 440, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 440  
; LENGTH: 2239  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-854-133-440

Alignment Scores:  
Pred. No.: 0.0022 Length: 2239  
Score: 523.00 Matches: 97  
Percent Similarity: 84.35% Conservative: 0  
Best Local Similarity: 84.35% Mismatches: 0  
Query Match: 98.31% Indels: 18  
DB: 9 Gaps: 1  
US-09-854-133-586 (1-97) x US-09-854-133-440 (1-2239)

QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
Db 2 GAGGTTGAAGTGACGACAGATCATGCCAGCTGGTGACAGTGAGACTGTCTCAACA 61  
QY 21 GluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 40  
Db 62 GAATTAAGGAAAAAGAAAGAAAGAAAGAGAGAGAGAGAAATCCAGGCCAATTGTGGC 121  
QY 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisTrpIleGln 60  
Db 122 ATAGATTTTATCATATTTTGGATTTTGGATTTTGGATTTTGGATTTTGGATTTTGG 181  
QY 61 GluSerLeuLeuCysProSerProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
Db 182 GAAAGCCTGTTGTGTCACCATCTCCAAAGGAGGTACCTGCAGGGAATGTTAAGGGA 241  
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArg 93  
Db 242 GGCTGCCCTTCCTGGGCAACAAGGAGGACCTGGGCAAGGAAAGTGCAGC 301  
QY 94 -----ArgLysCysSer 97  
Db 302 CCITTCAGGAAGAGACGCCTTTTCAGGAAGAGAGAAAGTGCAGC 346  
RESULT 4  
US-09-738-973-440  
; Sequence 440, Application US/09738973  
; Patent No. US20020110563A1









235 GGGCAGGAGAAAGTGCAC 2

235 GGGCAGGAGAAAGTGCAC 2



RESULT 13  
US-09-919-497-9  
; Sequence 9, Application US/09919497  
; Patent No. US20020106662A1  
; GENERAL INFORMATION:  
; APPLICANT: Mutter, George L.  
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER  
; FILE REFERENCE: B0801/7225  
; CURRENT APPLICATION NUMBER: US/09/919,497  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/221,735  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 9  
; LENGTH: 3995  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-919-497-9

Alignment Scores:  
Pred. No.: 1.19e+05 Length: 3995  
Score: 292.50 Matches: 69  
Percent Similarity: 24.53% Conservative: 9  
Best Local Similarity: 21.70% Mismatches: 11  
Query Match: 54.98% Indels: 229  
DB: 10 Gaps: 68

US-09-854-133-586 (1-97) x US-09-919-497-9 (1-3995)  
QY 1 GluVal-----Glu-----Val---SerArg-----Asp--- 7  
DB 2708 GAGATAGATGGCGGGATTACCACTTGTGTCTGTCCTCCGGGAGAAATGGAGAACATT 2767  
QY 8 -----His-----Ala-----Ser---Leu---Gly----- 12  
DB 2768 CAGCGGCACAACTTCACTTGGCGCGGCGCAGTACAAACAGCCACCTCTATGGGACCGTC 2827  
QY 13 AspSer-----Glu---Thr-----Leu-----Ser----- 18  
DB 2828 CAGTCCGTCGAGAGAGGTGGCAGAGCAGGGGAGAGCACTGCATCCTCGATGTCCTGGCCAAT 2887  
QY 19 -----Gln---ThrGluLeu-----Arg---Lys 24  
DB 2888 GCCGTGGCGGCTGCAGGCGGCGCCACCTGCACCCCATCGCCATCTTCACTCGGCCCGCG 2947  
QY 25 Lys---GluArg-----Lys---LysArg-----Glu-----ArgLys--- 34  
DB 2948 TCCCTGGAGAATGTGCTAGAGATTAAACAGCGGATCACAGAGGAGCAAGCCCGCAAGCC 3007  
QY 35 Phe-----Gln-----AlaAsnCys----- 39  
DB 3008 TTCGACAGAGCCACCAAGCTGGAGCAGGAGTTTCACAGAGTGTCTTCAAGCCATCGTGAG 3067  
QY 40 GlyIleAsp---Phe-----Ile-----Ile----- 45  
DB 3068 GGT---GACAGCTTTGAGGAGATCTACCACAAAGGTGACCGTGTCTATCGAGGACCTCTCA 3124  
QY 46 -----Phe---TrpIle-----Phe-----TrpIle---Leu 52  
DB 3125 GGCCTTACATCTGGGTTCCAGCCCGAGAGAGACTCTGATCCTGCTGCTTGGCTTGGCCTG 3184  
QY 53 -----LeuPheSerHisHis-----TrpIleGlnGluSer---LeuLeu 64  
DB 3185 GACTCGCCCTGCCTC-----CATCACCTGGGCCCTTGG-----TCTGGACTGAAT 3229  
QY 65 Cys-----ProPro---Ser-----Pro----- 69  
DB 3230 TGCCCAAGCCCTTGGCTCCCGCGGCTCCCTCCACCCCTTCTTATTATTCTTCTTCT 3289  
QY 70 Lys-----Glu---Val-----Thr-----Cys-----Arg-----Glu 76  
DB 76 Lys-----Glu---Val-----Thr-----Cys-----Arg-----Glu 76

Db 3290 AACTGGATCCAGCCTGTGGAGGGGGGACACTCTCTGATGTATCCCGCACCCAGAA 3349  
QY 77 MetLeu-----Thr-----GlyGly-----Cys--- 82  
DB 3350 ---CTGGCTCCTGAACGCCAGAACCTGGGGTCTGGGGGGAGCTGGGCTCCTTGTTC 3406  
QY 83 -----Leu-----ProTrpAla----- 86  
DB 3407 GAGCCCTTGTCTCTTAGGATCCCC---GCCCCCACCCTGCCCAATGCACACACAGACC 3463  
QY 87 -----Thr-----ArgSer----- 89  
DB 3464 ACCGGGGGCCACCTGCCCTCCCCCATCTCTCCACACACATTCAGAGAGTCAGGGCCCC 3523  
QY 90 -----His---Leu-----Gly---Arg-----Arg--- 94  
DB 3524 CTCGAGGAGCACCCGCTGCAGGATGTCAGGGGCCACAGGCTCCGCTCTCTCTAAGGCAG 3583  
QY 95 -----Lys-----Cys-----Ser 97  
DB 3584 GGTCTGGGTCAACCCCTGCCTCATCGTAATTCCTCATGTTACCTGATTCTCA 3637

RESULT 14  
US-09-254-783A-2  
; Sequence 2, Application US/09254783A  
; Patent No. US20020035734A1  
; GENERAL INFORMATION:  
; APPLICANT: Communil, Didier  
; APPLICANT: Boeynaems, Jeanmarie  
; TITLE OF INVENTION: G protein Coupled Receptor Showing Selective Affinity for  
; FILE REFERENCE: 9409/2062  
; CURRENT APPLICATION NUMBER: US/09/254,783A  
; CURRENT FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: PCT/BE98/00108  
; PRIOR FILING DATE: 1998-07-09  
; PRIOR APPLICATION NUMBER: EP97870101.9  
; PRIOR FILING DATE: 1997-07-09  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 2427  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-254-783A-2

Alignment Scores:  
Pred. No.: 3.13e+04 Length: 2427  
Score: 287.50 Matches: 63  
Percent Similarity: 30.36% Conservative: 12  
Best Local Similarity: 25.51% Mismatches: 17  
Query Match: 54.04% Indels: 155  
DB: 10 Gaps: 56

US-09-854-133-586 (1-97) x US-09-254-783A-2 (1-2427)  
QY 1 GluValGlu-----Val---SerArgAspHisAlaSer-----LeuGly----- 12  
DB 1120 GAA---GAGGCTCTGGGATCCCTTCAAGGACG---GCGAGCCTGGAGTGGTGAGGAC 1173  
QY 13 ---Asp---SerGlu-----ThrLeuSer-----Gln----- 19  
DB 1174 GATGTAACAGGAGAGATGATGATCATCGAGTATTCTGCCAGGCGGTGGCGAGGCGCC 1233  
QY 20 ThrGlu---Leu-----Arg-----Lys-----LysGlu-----ArgLys 28  
DB 1234 AGTGAGGACCTGTTCCCGAGGCCAAGCAAGAGCGGTTGCCAAGTCTCCAGGCGGGAAG 1293  
QY 29 LysLysArg---Glu-----Arg-----Lys-----Phe-----Gln 36  
DB 1294 CGGAAGCGGTGGGAAATGATCGAGGTGCCAAGTCTGCCCTGCCACTTCTTGGCAGCT 1353  
QY 37 Ala---Asn-----CysGlyIle-----AspPheIleIlePheTrp---Ile----- 48  
DB 48 Lys-----Glu---Val-----Thr-----Cys-----Arg-----Glu 48

```
Db 1354 GCCGACGACAACTCAGTGGGTCCAGGGGACTTC---CTG---TGGCCCATACTGGTG 1407
QY 49 -----PheTrp-----Ile-----Leu-----Leu-----PheSer----- 55
Db 1408 GTTGAAGTTCCTGGTGGCGGTGGCCAGCAATGGCCCTGGCCCTGTACCGCTTCAGCATCCGG 1467
QY 56 ---HisHis---Trp-----IleGln---Glu---Ser---Leu 63
Db 1468 AAGCAGCGCCCATGGCACCCCGCGTGGTCTTCTGTCCAGCTGGCAGTCAGCGACCTG 1527
QY 64 LeuCys-----ProPro-----Ser-----ProLys----- 70
Db 1528 CTCTGGGCTGTACGCTGCCCGCTGGCCGCTACCTCTATCCCCCAAGCAGTGGCGC 1587
QY 71 -----GluValThrCysArg---Glu---MetLeu---Thr-----GlyGly 81
Db 1588 TATGGGAGCGCGGTGCCGCTGGAGCGCTTCTTTCACCTGCAACCTGCTGGGCAGC 1647
QY 82 -----Cys----- 82
Db 1648 GTCATCTTCATCACCTGCATCAGCTCAACCGCTACCTGGGCATCGTGCACCCCTTCTTC 1707
QY 83 -----Leu---Pro-----Trp----- 85
Db 1708 GCCCGAAGCCACCTGCGACCCCAAGCAGCGCTGGCGGTGGTGGTCTCTG 1767
QY 86 -----Ala-----Thr---Arg---SerHisLeuGlyArg--- 94
Db 1768 GCCGCCCTGTGGCCATGCCACACTCAGCTTCTCCACCTG---AAGAGGCGCGCGCAG 1824
QY 95 Lys-----CysSer 97
Db 1825 CAGGGGCGGGCAACTGCAGC 1845
```

## RESULT 15

```
US-10-152-058-2
; Sequence 2, Application US/10152058
; Patent No. US20020142988A1
; GENERAL INFORMATION:
; APPLICANT: Communi, Didier
; TITLE OF INVENTION: G Protein Coupled Receptor Showing Selective Affinity for ATP
; FILE REFERENCE: 9409/2062
; CURRENT APPLICATION NUMBER: US/10/152,058
; CURRENT FILING DATE: 2002-05-21
; PRIOR APPLICATION NUMBER: PCT/BE98/00108
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: EP97870101.9
; PRIOR FILING DATE: 1997-07-09
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 2427
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-152-058-2
```

## Alignment Scores:

```
Pred. No.: 3.13e+04 Length: 2427
Score: 287.50 Matches: 63
Percent Similarity: 30.36% Conservative: 12
Best Local Similarity: 25.51% Mismatches: 17
Query Match: 54.04% Indels: 155
DB: 12 Gaps: 56
```

```
US-09-854-133-586 (1-97) x US-10-152-058-2 (1-2427)
```

```
QY 1 GluValGlu-----Val---SerArgAspHisAlaSer-----LeuGly----- 12
Db 1120 GAA---GAGGCCTCTGGGATCCCTTCAAGGACG---GCGAGCCTGGAGTGGGTGAGGAC 1173
QY 13 ---Asp---SerGlu-----ThrLeuSer-----Gln----- 19
||| ||| |||
```

```
Db 1174 GATGATGAACAGGAAGATGATGACATCGAGTATTTCTGCCAGGCGGTGGCGAGCGCCC 1233
QY 20 ThrGlu---Leu-----Arg---Lys-----LysGlu-----ArgLys 28
Db 1234 AGTGAGGACCTGTTCCCGAGGCCAAGCAGAAACGGCTTGCCAAAGTCTCCAGGGCGGAAG 1293
QY 29 LysLysArg---Glu-----Arg-----Lys-----Phe-----Gln 36
Db 1294 CGGAAGCGGTGGAAATGGATCGAGGTGCCAAGTCTGCTGCCCTGCCAACTTCTTGGCAGCT 1353
QY 37 Ala---Asn-----CysGlyIle-----AspPheIleIlePheTrp---Ile----- 48
Db 1354 GCCGACGACAACTCAGTGGGTTCAGGGGGACTTC---CTG---TGGCCCATACTGGTG 1407
QY 49 -----PheTrp-----Ile-----Leu---Leu-----PheSer----- 55
Db 1408 GTTGAAGTTCCTGGTGGCGGTGGCCAGCAATGGCCTGGCCCTGTACCGCTTCAGCATCCGG 1467
QY 56 ---HisHis---Trp-----IleGln---Glu---Ser---Leu 63
Db 1468 AAGCAGCGCCCATGGCACCCCGCGTGGTCTTCTGTCCAGCTGGCAGTCAGCGACCTG 1527
QY 64 LeuCys-----ProPro-----Ser-----ProLys----- 70
Db 1528 CTCTGGGCTGTACGCTGCCCGCTGGCCGCTACCTCTATCCCCCAAGCAGTGGCGC 1587
QY 71 -----GluValThrCysArg---Glu---MetLeu---Thr-----GlyGly 81
Db 1588 TATGGGAGCGCGGTGCCGCTGGAGCGCTTCTTTCACCTGCAACCTGCTGGGCAGC 1647
QY 82 -----Cys----- 82
Db 1648 GTCATCTTCATCACCTGCATCAGCTCAACCGCTACCTGGGCATCGTGCACCCCTTCTTC 1707
QY 83 -----Leu---Pro-----Trp----- 85
Db 1708 GCCCGAAGCCACCTGCGACCCCAAGCAGCGCTGGCGGTGGTGGTCTCTG 1767
QY 86 -----Ala-----Thr---Arg---SerHisLeuGlyArg--- 94
Db 1768 GCCGCCCTGTGGCCATGCCACACTCAGCTTCTCCACCTG---AAGAGGCGCGCGCAG 1824
QY 95 Lys-----CysSer 97
Db 1825 CAGGGGCGGGCAACTGCAGC 1845
```

Search completed: May 11, 2003, 20:09:26  
Job time : 94 secs

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: May 11, 2003, 17:46:32 ; Search time 1056 Seconds  
(without alignments)  
1487.654 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDASGLDSETLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62  
Xgapop 0.0 , Xgapext 0.5  
Ygapop 0.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0  
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09854133/runat\_05052003\_174426\_1224/app\_query.fasta\_1.263  
-DB=EST -QFMT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=5000  
-USER=US09854133\_ECGN\_1.1.1456@runat\_05052003\_174426\_1224 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=0 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=0 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Result No.	Score	Query Match	Length	DB ID	Description
1	339.5	63.8	936	12	BG284503 602408645
2	334.5	62.9	1072	12	BG388107 602413070
3	307.5	57.8	1143	12	BG260656 602372222
4	305	57.3	557	17	AQ002318 CIT-HSP-2
5	299.5	56.3	987	12	BG248967 602361925
6	294.5	55.4	1300	12	BG286262 602383093
7	293.5	55.2	1049	13	BM478521 AGENCOURT
8	293	55.1	882	12	BF699490 602126079
9	292.5	55.0	910	12	BG326527 602425373
10	290	54.5	937	12	BG391278 602417323
11	288.5	54.2	1062	10	BE035217 MO01R03 M
12	287	53.9	862	14	BQ688218 AGENCOURT
13	287	53.9	876	14	BQ687348 AGENCOURT
14	286.5	53.9	748	17	AZ740600 RPCI-24-9
15	286.5	53.9	1021	12	BG261332 602373094
16	286.5	53.9	1051	12	BG026259 602291607
17	286	53.8	913	12	BF696158 602124671
18	285.5	53.7	960	17	CNS03MEM Tetraodon
19	285	53.6	597	17	BH787408 fzm014f0
20	285	53.6	836	10	BE538494 601068145
21	285	53.6	956	12	BG169397 602320964
22	285	53.6	998	12	BE793114 601583173
23	285	53.6	1119	17	CNS06QSG T3 end of
24	285	53.6	1210	13	BM561270 AGENCOURT
25	285	53.6	1242	17	B12259 F28B2-T7 IG
26	284.5	53.5	1006	14	BQ726437 AGENCOURT
27	284	53.4	1310	11	AF116717 Homo sapi
28	284	53.4	1780	12	BG115708 602317180
29	283.5	53.3	1092	17	CNS038E2 Tetraodon
30	283	53.2	880	10	BE570665 601329996
31	283	53.2	960	17	CNS041ML Tetraodon
32	283	53.2	989	17	CNS03D07 Tetraodon
33	283	53.2	1136	14	BM928444 AGENCOURT
34	282.5	53.1	936	13	B1199098 602759269
35	282.5	53.1	1038	12	BE913378 601665590
36	282.5	53.1	1271	13	BI687475 603315028
37	282	53.0	839	10	BE641384 Cri2_3_C0
38	282	53.0	920	13	BM465182 AGENCOURT
39	282	53.0	943	12	BG536746 602564875
40	282	53.0	1769	12	BG170413 602322745
41	282	53.0	2917	11	AK019171 Mus muscu
42	281.5	52.9	835	12	BG253649 602362916
43	281.5	52.9	853	17	AZ740605 RPCI-24-9
44	281.5	52.9	886	12	BF215972 601881522
45	281.5	52.9	1023	12	BE797979 601584187

ALIGNMENTS

RESULT 1  
BG284503  
LOCUS BG284503 936 bp mRNA linear EST 21-FEB-2001  
DEFINITION 602408645F1 NIH\_MGC\_91 Homo sapiens cDNA clone IMAGE:4537810 5',  
rna sequence.  
ACCESSION BG284503  
VERSION BG284503.1 GI:13035516  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 936)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10463 row: i column: 11  
High quality sequence stop: 795.  
Location/Qualifiers  
1. .936  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4537810"  
/clone\_lib="NIH\_MGC\_91"  
/tissue\_type="adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: prostate; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.4 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."  
BASE COUNT 248 a 202 c 251 g 235 t  
ORIGIN

FEATURES  
source

Alignment Scores:  
Pred. No.: 356 Length: 936  
Score: 339.50 Matches: 62  
Percent Similarity: 57.25% Conservativeness: 13  
Best Local Similarity: 47.33% Mismatches: 7  
Query Match: 63.82% Indels: 49  
DB: 12 Gaps: 28  
US-09-854-133-586 (1-97) x BG284503 (1-936)  
QY 1 GluValGluValSerArg---AspHis---Ala-----Ser-----Leu 11  
Db 41 GAG---GAGTGGAGAAATTGAGAGCAGGATGATACACAGGTGTTCTGAGTAGTAATTA 97  
QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22  
Db 98 ---GATCGCTGTGAAGGAAACACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 151  
QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33  
Db 152 CGCTGAGAGAGACAGTCTGAAGCAGGAGGAGATCATCATAGTACACCAAGAGACACC 211  
QY 34 LysPheGlnAlaAsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---Phe 49  
Db 212 AAA-----GTTGAAAGTTT---GTTTCTTTCCTCTCTGTTTATT 250  
QY 50 Trp-----Ile---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysPro 66  
Db 251 TTCCCGCGTGTCTCCCTACTA-----TGG---TCAGAAAGCCTGTGTGTCCA 295  
QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla 86  
Db 296 CCATCTCCAAAGGAGGTACCTGCAGGGAATGTTAACGGAGGCTGCCTTCCCTGGGCA 355  
QY 87 ThrArgSerHisLeuGlyArgArgLysCysSer 97  
Db 356 ACAAGGAGCCACCTGGGCAGGAGAAAGTGCAGC 388

RESULT 2  
BG388107  
LOCUS  
DEFINITION 602413070F1 NIH\_MGC\_92 Homo sapiens CDNA clone IMAGE:4521736 5',  
mRNA sequence.  
ACCESSION BG388107  
VERSION BG388107.1 GI:13281553

KEYWORDS  
SOURCE

EST.  
human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1072)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10421 row: k column: 17  
High quality sequence stop: 625.  
Location/Qualifiers  
1. .1072  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4521736"  
/clone\_lib="NIH\_MGC\_92"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."  
BASE COUNT 331 a 229 c 270 g 242 t  
ORIGIN

FEATURES  
source

Alignment Scores:  
Pred. No.: 990 Length: 1072  
Score: 334.50 Matches: 62  
Percent Similarity: 57.36% Conservativeness: 12  
Best Local Similarity: 48.06% Mismatches: 10  
Query Match: 62.88% Indels: 45  
DB: 12 Gaps: 28  
US-09-854-133-586 (1-97) x BG388107 (1-1072)  
QY 1 GluValGluValSerArg---AspHis---Ala-----Ser-----Leu 11  
Db 59 GAG---GAGTGGAGAAATTGAGAGCAGGATGATACACAGGTGTTCTGAGTAGTAATTA 115  
QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22  
Db 116 ---GATCGCTGTGAAGGAAACACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 169  
QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33  
Db 170 CGCTGAGAGAGACAGTCTGAAGCAGGAGGAGATCATCATAGTACACCAAGAGACACC 229  
QY 34 LysPheGlnAlaAsnCysGlyIleAspPhe-----IleIlePheTrpIlePhe--- 49  
Db 230 AAAGTTGAA---AGTTTA---GTT---TTCCTTCCCTCTGTTTATT---TTCCTC 274  
QY 50 TrpIle---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysProSer 68  
Db 275 CGTGTGTCTCCCTACTA-----TGG---TCAGAAAGCCTGTGTGTCCACCATCT 319  
QY 69 ProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAlaThrArg 88  
Db 320 CCAAAGGAGGTACCTGCAGGGAATGTTAACGGAGGCTGCCTTCCCTGGCAACAGG 379  
QY 89 SerHisLeuGlyArgArgLysCysSer 97  
Db 380 AGCCACCTGGGCAGGAGAAAGTGCAGC 406





FEATURES  
source

Location/Qualifiers  
1. .557  
/organism="Homo sapiens"  
/db\_xref="GDB:7147907"  
/db\_xref="taxon:9606"  
/clone="2283E8"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelOBAC11; site\_1: HindIII; site\_2: HindIII"  
HindIII 142 a 152 c 133 g 130 t

BASE COUNT 142 a 152 c 133 g 130 t  
ORIGIN

Alignment Scores:

Pred. No.: 551 Length: 557  
Score: 305.00 Matches: 55  
Percent Similarity: 57.52% Conservative: 10  
Best Local Similarity: 48.67% Mismatches: 10  
Query Match: 57.33% Indels: 38  
DB: 17 Gaps: 22

US-09-854-133-586 (1-97) x AQ002318 (1-557)

QY 7 AspHisAlaSerLeuGlyAspSerGluThr---LeuSer---Gln-----ThrGluLeu 22  
Db 557 GATCACTGTGAA---GGATATAGC---ACACCTTTGAGTTTTCACCTGTGAACA---CTA 507  
QY 23 ---Arg---Lys---Glu---Arg---Lys---LysLys-----ArgGlu 32  
Db 506 TAGCGTGAGAGATACAGTCTGAAGCAGAGTAAGACATGCATCAGTAACACCAAGAGAC 447  
QY 33 ArgLysPheGlnAlaAsnGlyGlyIleAsp---PheIleIlePheTrp-----Ile 48  
Db 446 ACCAAA-----GTTGAAAGTTT---GTTTCTTTCCTCTGATTTA 408  
QY 49 PheTrpIleLeuLeuPheSerHisHis-----TTPileGlnGluSerLeuLeu 64  
Db 407 TTT-----TTC---CACCGTGTGTCCTACTATGG---TCAGAAAGCCTGTTG 366  
QY 65 CysProProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuPro 84  
Db 365 TGTCCACCATCTCCAAAGTAGTTTACCTGCAGGAAATGTTAACGGAGGCTGCCTTCCC 306  
QY 85 TrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
Db 305 TGGGCAACAAGGAGGCCACCTGGGCAGCAGAAAGTGCAGC 267

RESULT 5  
BG248967  
LOCUS

DEFINITION BG248967 987 bp mRNA linear EST 13-FEB-2001  
602361925F1 NIH\_MGC\_89 Homo sapiens CDNA clone IMAGE:4470054 5',  
mRNA sequence.

ACCESSION BG248967  
VERSION  
KEYWORDS  
SOURCE EST. GI:12758783

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 987)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

REFERENCE

AUTHORS  
TITLE  
JOURNAL  
COMMENT

Plate: LLAM10287 row: b column: 07  
High quality sequence stop: 493.  
Location/Qualifiers  
1. .987  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4470054"  
/clone\_lib="NIH\_MGC\_89"  
/tissue\_type="hypernephroma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
BASE COUNT 274 a 299 c 229 g 185 t  
ORIGIN

Alignment Scores:

Pred. No.: 1.16e+04 Length: 987  
Score: 299.50 Matches: 64  
Percent Similarity: 28.99% Conservative: 16  
Best Local Similarity: 23.19% Mismatches: 9  
Query Match: 56.30% Indels: 187  
DB: 12 Gaps: 56

US-09-854-133-586 (1-97) x BG248967 (1-987)

QY 1 GluValGlu-----ValSerArgAspHisAla---SerLeu-----Gly---AspSer 14  
Db 64 GAGGTGGAGGTTGCAGTGAGCCGAGATCATGCCACTGCACCTCCAGCTGGGCAACAGAGT 123  
QY 15 GluThrLeuSerGlnThrGluLeuArgLysLysGluArgLys-----Lys--- 29  
Db 124 GAGACTTTGTCTCAAGGAAA---AAAAAATAAATAAATCTGGGGATGAAATAA 180  
QY 30 ---Lys---Arg---Glu-----ArgLys-----Phe 35  
Db 181 CCAAAATGAAATAGCTAGAAACTCAGCAAGCAGAAAGCGTCCCTTTTCTCACCCCTTTT 240  
QY 36 -----Gln-----Ala----- 37  
Db 241 GGTCCCTGGCGGATAAAATCAGTCACCTATACAAACAACTGAAAGACGCTCTGCTTTAA 300  
QY 38 Asn-----Cys-----GlyIleAspPheIle----- 44  
Db 301 CACCAATGATGCACAGCAGTACTTACTTATGTATTCGAGGT---GACCTTATATAGATG 357  
QY 45 -----IlePheTrp-----Ile---Phe-----Trp-----Ile 51  
Db 358 GAGAAGAGGCTGCATTGGCGAGAACTGATGTATAGGAACCCCATTTGGTTCAAGCTTCTC 417  
QY 52 -----Leu-----LeuPheSerHis-----His 57  
Db 418 CCTGCGAGGACAAATGCTCTCGGTTCATTATAGCTCTTTAC---CACAGACTGGGCCAT 474  
QY 58 -----Trp----- 58  
Db 475 TCTAGTGAACCCGGTGGGTACCCCTTGGGGTTGCCAAGAAAACCATTTCTTATCTGGTTT 534  
QY 59 ---Ile---Gln-----Glu-----Ser---LeuLeuCys----- 65  
Db 535 CCACTTACCTCAGGAGCGGAGCAGATAACGGGGTTCATCCAAATC---TGTATTGCAACG 591  
QY 66 -----Pro---ProSer---Pro---LysGluValThrCysArgGluMet 77  
Db 592 CTATCTCATCTCTCCCAACCTTCTCTCTCTGGGAAACCC---AACTGC-----CTA 642  
QY 78 -----LeuThr---Gly-----Cys---Leu-----Pro--- 84  
Db 643 TCTCATCTGACCAAAAGGTCTAGCAGGATACGCCTGACACGACTCGGCAGGCACCCCTAT 702  
QY 85 -----Trp-----Ala-----Thr----- 87

```
Db 703 CGGCGCTGGGCGGAAGCCCTTGCCTAAACCCAGAGAGCCGGCGGACTAAACAAGGG 762
QY 88 Arg-----Ser---His-----Leu---Gly 92
Db 763 AGGCCCAACAGGGCGGGCGGCATCCAGCACTCCAGGCCCGCCCGGGAACGGACGGT 822
QY 93 -----ArgArg-----Lys-----Cys 96
Db 823 TGCACACACCTGGCTAAACGGAGGCGCTATCCAGCCTCTTCTCCCTGC 870

RESULT 6
BG286262 1300 bp mRNA linear EST 21-FEB-2001
LOCUS 60238303F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4500599 5',
DEFINITION mRNA sequence.
ACCESSION BG286262
VERSION BG286262.1 GI:13038979
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1300)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10366 row: j column: 24
High quality sequence stop: 366.
Location/Qualifiers
1. 1300
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4500599"
/clone_lib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
1 others

BASE COUNT 287 a 390 c 313 g 309 t
ORIGIN

Alignment Scores:
Pred. No.: 6.08e+04 Length: 1300
Score: 294.50 Matches: 66
Percent Similarity: 29.89% Conservative: 12
Best Local Similarity: 25.29% Mismatches: 12
Query Match: 55.36% Indels: 171
DB: 12 Gaps: 50

US-09-854-133-586 (1-97) x BG286262 (1-1300)
QY 1 Glu-----ValGluValSer---ArgAspHisAla---SerLeu-----GlyAsp--- 13
Db 180 GAGCGGGGGTTCAGTGAGCTAAAT---CATGCCACTGCATTCACACCTGGGAACAG 236
QY 14 SerGluThrLeuSerGlnThrGluLeuArgLysLysGluArgLysLys---Lys----- 30
Db 237 AGTGAGACCCCTGTCTCAA---AAA---AAAAAAAAAAAAAAAAAAATCCAAATTCGGG 290
QY 31 ---ArgGluArg-----Lys---PheGlnAlaAsnCys---GlyLeuAspPhe 43
```

```
Db 291 GTGCGGAAAGATGGGAAAAAACCCTTT---TTGGGGTGTGATGGGACCCCATTT 347
QY 44 Ile-----Ile---Phe-----TrpIlePheTrp-----IleLeu 52
Db 348 GTGTTGAAAAATTTTCTCCCGTGCAGGTGCTGG-----TGGTCCCACCCATCTT 401
QY 53 -----Leu-----PheSerHis-----HisTrp----- 58
Db 402 ATTTACCCCTCCNCTATTTCTCTCATTTCTCATCTCATTTTCTTATAAAC 461
QY 59 -----Ile-----Gln----- 60
Db 462 ACCCTTATACCTTATGTGGTGTATAGGCGCACAGGTTCTCTTAAAGAGACCTTGGAG 521
QY 61 -----Glu-----Ser---Leu---LeuCys----- 65
Db 522 AGAATCCCCCGGGGAGAGAACTAAGAAGCCCCCTGTGTGTGTCTCTCCGTGTGGAGCC 581
QY 66 Pro-----Pro---Ser---ProLysGlu--- 71
Db 582 CCTTTGTTTGTCTTCTTAGCCGGGGGGGGGCACACCCACCTGGGAGACCC---GATTTT 638
QY 72 -----Val---Thr-----Cys----- 74
Db 639 ACACCTTTTGTGTTTCAACCTTCTGTGTGTCTCCGCTCTGTGTCTCTCACACGGGCGCTT 698
QY 75 ---ArgGluMet---Leu---Thr----- 79
Db 699 CTCGCGAGGGTGGCCCTTCTCGCCCAACCGGCTGTTTTAGCCACACGGGGGGGCA 758
QY 80 -----Gly-----Gly-----CysLeu 83
Db 759 CCGCACGCGGGCGCAACTCGTGGCGTCTCTCTCGGGGGCGCCCGCGTGTGTCTC 818
QY 84 ---Pro---Trp-----Ala---Thr-----ArgSerHisLeu----- 91
Db 819 CCCCCGAGATGCTGGGCCATCTGCGGCATTACACACCTCTGCGCTCTATTTGTAGCAC 878
QY 92 -----GlyArg-----Arg-----LysCys 96
Db 879 CCCCAGGGAGATCTTCTCCGAGACAAATTTCTTTGGTGGACCGCCCGCATCTGT 938
QY 97 Ser 97
Db 939 GCC 941

RESULT 7
BM478521 1049 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6457605 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5575203
DEFINITION 5', mRNA sequence.
ACCESSION BM478521
VERSION BM478521.1 GI:18527563
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1049)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12325 row: b column: 04
```

FEATURES  
source High quality sequence stop: 705.  
Location/Qualifiers  
1. 1049  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5575203"  
/clone\_lib="NIH\_MGC\_92"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: testis; Vector: pCMV-Sport6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
BASE COUNT 286 a 253 c 240 g 262 t 8 others  
ORIGIN

Alignment Scores:  
Pred. No.: 2.48e+04 Length: 1049  
Score: 293.50 Matches: 67  
Percent Similarity: 27.84% Conservative: 14  
Best Local Similarity: 23.02% Mismatches: 11  
Query Match: 55.17% Indels: 199  
DB: 13 Gaps: 61

US-09-854-133-586 (1-97) x BM478521 (1-1049)

QY 1 GluValGlu-----ValSerArg-----Asp----- 7  
Db 186 GAGGTGGAGGGAAGATGGTCAGTAGGACAGAGGTAACATTGATGACTCGCTCATGGT 245

QY 8 ---HisAlaSerLeu-----GlyAsp---SerGlu-----Thr--- 16  
Db 246 GGAATGCTCCGCTGAAGGCCCGGAGGAGGTACCGAAGACACAGTAATCACTGGT 305

QY 17 -----LeuSerGln-----Thr---Glu---Leu 22  
Db 306 GTGCATATTGTCATGAACCATCCTG---CAGGAACAAGTTTCACAAAGAACGCTAC 362

QY 23 ArgLys-----LysGlu---ArgLys-----Lys---Lys-----Arg--- 31  
Db 363 AAGAAGTACATCAAGATTACATGAATCAATCAAGGGAAGTACCGAAGACAGAGACCA 422

QY 32 GluArg---Lys---Phe-----Gln-----Ala--- 37  
Db 423 GAAAGAGTAAACCTTTATGACAGGGGCTGCAGAACAAATCAAGCACATCCTTGCTAAT 482

QY 38 -----AsnCys-----Gly---Ile----- 41  
Db 483 TTCAAAAACATACAGTCTTTTATTGGTGAACACATGAATCCAGATGGCATGTTGCTCTA 542

QY 42 ---AspPhe-----Ile-----Ile---Phe----- 46  
Db 543 TTGGACTACCGTGAGGATGGTGACCCCATATATGATTTCTTTAAGGATGTTTAGAA 602

QY 47 -----Trp---IlePheTrpIle-----Leu-----Leu 53  
Db 603 ATGGAATAATGTTAACAATGTGGCAATATTGATCTATCATCCTGTCATCATAACTG 662

QY 54 PheSer-----His---His-----Trp---IleGlnGluSer 62  
Db 663 GCTTCTGCTGTGTCATCCACACACACACAGGACTTAAGACAAATGGGACTGATGTCATCT 722

QY 63 -----Leu-----Leu-----Cys----- 65  
Db 723 TGAGCTCTTCATTTATTGACTGTGATNTTATTGGAGTGAGGAGCATTTTAAAGAA 782

QY 66 -----Pro-----Pro-----Ser----- 68  
Db 783 AACATGCCATGTAGTGGCCCTAAATAAATGCAATTAACCTCCNNNNNNAAAAACCC 842

QY 69 -----ProLys-----GluValThrCys---Arg---GluMet----- 77  
Db 69 -----ProLys-----GluValThrCys---Arg---GluMet----- 77

Db 843 CCCACATNCCAAACATTTCTCA---ACGGGGGGCGCTCTAAATATCCCTCGGAGG 899

QY 78 -----LeuThr-----Gly---GlyCys-----Leu---Pro---Trp 85  
Db 900 GGTCCAGCTTACCCCGTAACCGGCTTTCTTCTTAACAAGGTGGGCCCTATTATTGG 959

QY 86 -----Ala---Thr-----Arg--- 88  
Db 960 GGCCGGTTATTACCGCTTAGGTACAGTCAGCGGCTCTTTTATACAAACCGGCTCGTGA 1019

QY 89 SerHisLeuGlyArgArgLys---Cys---Ser 97  
Db 1020 TCT-----GGG---GGAAAAACATGCCTAACCC 1043

RESULT 8  
BF699490 882 bp mRNA linear EST 22-DEC-2000  
LOCUS 602126079F1 NIH\_MGC\_56 Homo sapiens cDNA clone IMAGE:4282962 5',  
DEFINITION mRNA sequence.  
ACCESSION BF699490  
VERSION BF699490.1 GI:11984898  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 882)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: CLONETECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1113 row: n column: 19  
High quality sequence stop: 672.

FEATURES  
source Location/Qualifiers  
1. 882  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4282962"  
/clone\_lib="NIH\_MGC\_56"  
/tissue\_type="primitive neuroectoderm"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggccgcctcgcc); Site\_2: SfiI (ggccattatggcc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-ATTCTAGAGCGGAGCGGCGGACATG-dT(30)BN-3,  
sequence: 5'-ATTCTAGAGCGGAGCGGCGGACATG-dT(30)BN-3,  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."

BASE COUNT 282 a 167 c 222 g 211 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.18e+04 Length: 882  
Score: 293.00 Matches: 67  
Percent Similarity: 28.83% Conservative: 12  
Best Local Similarity: 24.45% Mismatches: 6  
Query Match: 55.08% Indels: 189  
DB: 12 Gaps: 56

US-09-854-133-586 (1-97) x BF699490 (1-882)



us-09-854-133-586.rst

Tue May 13 12:12:51 2003

```

QY 1 Glu---ValGlu---Val-----Ser-----ArgAspHisAla----- 9
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37 GAAGCAGTTGAGTGTGCAGCGGCGAGTCGTCGCCCTGCGGC---CGTGCCACCAAGGA 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 10 SerLeu-----Gly-----AspSerGluThr-----LeuSerGln 19
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 AGCATCCAGACACCGGGAAGATGGAAGATGGT---ACCCAAAGCATATCATC---CAG 147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 20 ---Thr-----Glu---Leu---Arg-----Lys----- 24
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148 ATGACAGGATTAAAGATGGAAGAAAGAGCGCTAGTCAAAATTAATTTAAACTAGAT 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 25 -----Lys---GluArgLys---Lys-----LysArg 31
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208 TGCACCTTTTAAAGAGTGAG---AAATACAAAATTTGTACACATCTTATAGCTGAACGC 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 32 -----GluArgLysPhe---GlnAlaasnCys-----Gly-----Ile--- 41
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 CTATGTAAGAGTGAA---AAATTTTAGCAGCT---TGTCGGCAGGAAAGTGGATACTA 318
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 42 -----AspPheIlellePhe-----TrpIle-----Phe 49
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 319 ACCAAGGACTATATAATTTCATAGTGGCCAAAGTGGCAGATGGCTTGATGAAACAACTAT 378
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 50 ---Trp---Ile-----Leu-----LeuPhe---SerHis----- 56
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 379 GAATGGGGGATATAACCACTTGAAACAAAGATTCCCGTTATTCACCTCAATGCAATCTGC 438
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 57 -----HisTrp-----Ile---Gln--- 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 439 ACCTAAAGATGGCGTGAAGAACTGAAACGCACTGGTGTCCAGGAGCCTTCCACAGATG 498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 GluSerLeuLeuCysProPro-----Ser 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 499 GAAAGT-----TGTCCTCCTTGTTAGAACTGATAAGCGAAGTGATTCTCTTATAAGAGT 552
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 69 -----ProLys-----GluVal---ThrCy 74
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 553 TTGGAGGCTGGAAGGCAAAATGTTATTTTACCAAAAAGTTCCACAGTGGAAATACTCATG 612
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 74 s-----Arg-----Glu-----MetLeu----- 78
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 613 TGATGGCAGTAATGCAAGAATTACGGCTGAGAAGAAAGAAAGATTAAGGCTCAATTAT 672
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 79 -----Thr-----GlyGly----- 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 673 CAATCAGACCCAGGGGATTTCTTTTGAAGAGGAATCCGAGGTGGAGATCCCAACCAATT 732
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 82 -----CysLeuProTrpAlaThrArg-----SerHisLeuGly----- 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 733 GTTGGCGTACTGT---CCGTGGAACACAAAGATTCGGAGAGTCG-----GGTCTGTAGAA 783
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 93 -----Arg-----Lys-----Cys 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 784 GGCTGAACGTTTAAACGAGAAAGACTAAGTTTGGGTTTTCG 823
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 9
BG326527
LOCUS 602425373F1 NIH_MGC_14 Homo sapiens cdna clone IMAGE:4562994 5',
DEFINITION mRNA sequence.
ACCESSION BG326527
VERSION BG326527.1 GI:13132964
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 910)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

```

```

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
CDNA Sequencing by: Incyte Genomics, Inc.
CDNA distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1276 row: b column: 19
High quality sequence stop: 706.
Location/Qualifiers
1..910
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4562994"
/clone_lib="NIH_MGC_14"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 237 a 178 c 249 g 246 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1.41e+04
Score: 292.50
Percent Similarity: 57.66%
Best Local Similarity: 49.55%
Query Match: 54.98%
DB: 12
Length: 910
Matches: 55
Conservative: 9
Mismatch: 8
Indels: 40
Gaps: 21

```

US-09-854-133-586 (1-97) x BG326527 (1-910)

```

QY 6 ArgAspHisAlaSerLeuGlyAsp---Ser-----GluThr-----LeuSerGlnThr 20
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 AAG---CAC---ACCTTT---GAGTTTTCACCTGTGAACACTATAGCGCTG-----ACA 50
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 21 GluLeuArgLys-----LysGluArgLys-----Lys-----LysArg---Glu 32
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 GAG-----ACAGTCTGAAAGCAGAGGAGAGACATCGATCAGTAACACCAAGAGACACCAA 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 33 Arg---LysPheGlnAlaasnCysGlyIleAspPhe-----IleIlePheTrpIle 48
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 105 AGTGAAGAGTTT-----GTT---TTCTTCCCTCTGTGTTTATTT----- 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 49 Phe---TrpIle---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysPro 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 TTCCCCCGTGTGTCCCTACTA-----TGG-----TCAGAAAGCCTGTG-TGTCCA 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla 86
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 CCATCTCCAAAGGAGGTTACCTGCAGGAAATGTTAACGGGAGGCTGCCTTCCTGGGCA 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 87 ThrArgSerHisLeuGlyArgArgLysCysSer 97
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 ACAAGGAGCCACCTGGGCGAGGAGAAAGTGCAGC 277
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 10
BG391278
LOCUS 602417323F1 NIH_MGC_92 Homo sapiens cdna clone IMAGE:4536772 5',
DEFINITION mRNA sequence.
ACCESSION BG391278
VERSION BG391278.1 GI:13284726
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

EST 12-MAR-2001  
linear  
937 bp  
mrna  
Homo sapiens  
cdna clone  
IMAGE:4536772 5',

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 937)  
NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

state: LLAM10460 row: n column: 05  
High quality sequence stop: 699.  
Location:

## source

Location/Qualifiers  
1. .937

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone="IMAGE:4536772"  
/clone_lib="NIH_MGC_92"  
/tissue_type="embryonal c  
/lab_host="DH10B (phage
```

Site\_2: SaliI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NTH\_MGC Library." 202

[illegible]

235 a 202 c 249 g 251 t

Alignment Scores:

```
Pred. No.: 1.97e+04
Score: 290.00
Percent Similarity: 26.49%
Best Local Similarity: 22.85%
Query Match: 54.51%
DB: 12
```

US-09-854-133-586 (1-97) x BG391278 (1-937) Gaps

```

QY      1 GluValGlu-----ValSerArgAspHisAla---SerLeu-----GlyAsp---Ser 14
          |||||
          25 GAGGTGGAGGTGCAGTGAGCCGAGATCATGCCACTACACTTCAGCTGGCCCAATAGAGT 84
QY      15 GluThrLeuSer-----

```

Db	85	GAGACTCTGTCTCAAAAACAAACAAACAGAAAAAACAACAAAGAACCTAGT	20
			---Gln---Thr---
QY	21	-----Gln-----	

Db 145 TTGGGTTTGGCAGAAATCATCATCAGGAATATTACAGACCCCAGTAGGTAACCTTGAGTG 23  
 ---Glu-----Leu-----Arg-----  
 ||| ||| |||  
 QY 24 ---LysLysGlu----- 204

dbb  
205 CTCAGGTCCTGACCTGAGGTCTCAGTCTTTATGGCTTCATGAAGGACTGTGCCCTCAGG 264

**b** 265 GAGGGACCATGGCTTGGCTTGTGGGGTCTTGGTGATGACTGCCATCTTCTTCATCACC 32  
 |||  
**y** 33 ----- 324

b  
y

```
325 ACACCCAGCTTCTTGTGGCATTTGCGAGGACAGCAGACAGATAGAGGTTTACCCCTTT 35
      |||                               Lys  : : : Phe  |||
      -----Arg-----
36 -----Gln-----AlaAsnCys----- 384
```

385 ATCTCCCACTGAGCAAGATGCTTCCCTGTTCCAGAGGAGTAACACTGCTTATGCTTGA 444  
 ||| . ||| ||| Gly-Ile ---As 42  
 42 p-----PheIleIlephe-----TATGCTTGA 444  
 ||| ||| ||| ||| ::

```
|
|||
||||
-----TrpIle---phe---Trp-----IleLeu----Le 53
|||:: ||| ||| ||||| ||
```

1

1

Db	445	CTGCTGTTTT	---TTTTTGGTCTTTGGTTGGCTTTTCTTTGGATGGGAATTCGTATT	501
QY	53	u	-----PheSer---His-----His-----	
Db	502	AAGATGTTAAGTGTGTGCAGCTTACATATGGAACATCCAGGAGAGGTGACTGGCCCC	58	Trp
QY	59	-----IleGlnGluSer-----Leu-----		
Db	562	CACCCTGCTT---GGCAGTGGATTCTGCATCACCAGCTTCAGAAGTGT---CCCTGTTTC	68	Se
QY	68	r	-----ProLys-----GluValTh	73
Db	616	GCTAGTCTGTGGTTGCTGTGCTTGCCTAAATACTTGGGTCCCAAGCTGAGACAA	670	
QY	73	rCys	-----ArgGluMe	77
Db	671	TGTGCTGTGTAAACGGGAAGAGTCAATCTGTGAAGGGTATCATTTTCAAGAAAGAA	727	
QY	77	tLeuThrGly	-----Gly-----Cys-----LeuProTrp	85
Db	728	TTGGCCGGTGTGCCCTTCACGAGGAGTCCAGTCTGCGGTGGATTTTTTCACTGGGGATG	786	
QY	86		-----Ala-----Thr-----ArgSer	89
Db	787	ACGGAACGGGATTATTGGCCGGGGTTTTTCACGGGGTTGCCATACAATTTGGCCGATCTTG	846	
QY	90	-HisLeuGly	-----Arg-----Arg-----Lys-----Cys-----	96
Db	847	GCATTATGGGACCGAAAGAATACCCCGGTTAGAACTCTGTGTTGGGAACAAGAGGACC	906	
QY	97	-Ser	97	
Db	907	CTCT	910	
RESULT	11			
BE035217				
LOCUS	BE035217			
DEFINITION	MO01E03 MO Mesembryanthemum crystallinum cDNA 5', mRNA sequence.	1062 bp	mRNA	linear
ACCESSION	BE035217			
VERSION	BE035217			

VERSION	BE035217.1	GI:8330341
KEYWORDS	EST.	
SOURCE	common iceplant.	
ORGANISM	Mesembryanthemum crystallinum	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.	
AUTHORS	1 (bases 1 to 1062)	
	Bohnert, H.J., Borchert, C., Brazille, S., Brooks, J., Eaton, M., Ferrel, H., Kawasaki, S., McCollough, A., Michalowski, C.B., Palacio, C., Scara, G., Wheeler, M. and Zepeda, G.R.	
TITLE	Functional Genomics of Plant Stress Tolerance	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Michael	

Contact: MICHALOWSKI, C.B.  
University of Arizona  
Bio Sciences West room 513, Tucson, AZ 85721, USA  
Tel: 520-621-7982  
Fax: 520-621-1507

Email: cbm@u.arizona.edu  
 Best blastx match: 'gi|3236235 (AC004684) unknown protein  
 [Arabidopsis thaliana] gi. . . 116 2e-25'  
 Location:

```

source
location/Qualifiers
1. .1062
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone_library="

```

```

/clone_id="MO"
/tissue_type="apical meristem and leaf primordia"
/dev_stage="5 weeks"
/note="no stress"
319 a 185 0 0.0

```

100 c	245 g	310 t	2 others
<p>ignment Scores: ed. No.:</p>			
3.94e+04			

Length: 1062

1



LOCUS	AZ740600	748 bp	DNA	linear	GSS 25-JAN-2001
DEFINITION	RPCI-24-90HL3.TJ RPCI-24 Mus musculus genomic clone RPCI-24-90HL3, DNA sequence.				
ACCESSION	AZ740600				
VERSION	AZ740600.1 GI:12511838				
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 748)				
AUTHORS	Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P. and Fraser, C.M.				



**Mismatches:**  
**Indels:**

DB: 12 Gaps: 59

US-09-854-133-586 (1-97) x BG261332 (1-1021)

QY 1 GluValGlu---Val-----Ser-----ArgAsp-----His--- 8  
Db 11 GAG---GAGAAGGTACGTACAGCAAGCTGGGTAGCAGGAGATCCAAAGGATATCATGAA 67

QY 9 -----Ala-----SerLeu----- 11  
Db 68 GTTCCAGGGCCTTTGGAACACCAGAGATTGCTTCCTGTTGGAAAGGCAATCACTAG 127

QY 12 Gly-----Asp-----SerGlu-----Thr 16  
Db 128 GGAAGCACAGATGTGGAAGTGAATGTGCGGAAATGCCTTCAATCAGAAATGTTCTCC 187

QY 17 Leu-----Ser-----GlnThr-----Glu---Leu 22  
Db 188 ATCCAGAGAGATGAAGTAATTCATGCTGGCTAGCAGTCTTTGGATAGGTTTTAGTACCCTAAAGGC 307

QY 23 ---ArgLys-----Lys-----Glu-----ArgLys--- 28  
Db 248 CCCAGAAACATTTGCTCTGGCTAGCAGTCTTTGGATAGGTTTTAGTACCCTAAAGGC 307

QY 29 ---LysLys---ArgGlu-----ArgLys-----Phe-----GlnAlaAsn 38  
Db 308 TCATCCAAATACTTGAGTTGATGTGATGCAATCAGCTGTTTTTCTAGTGCCTAA---GAC 364

QY 39 Cys-----GlyIle-----Asp----- 42  
Db 365 TGTGAGGAAGATGAGAGAAATCCAGTACTAAAGGTATGGCAAGAGACAGTTTCTGTGG 424

QY 43 ---PheIleIle-----Phe-----Trp-----IlePheTrpIle-----Leu--- 52  
Db 425 ATGTTCTCATCTGAAATTTTGAGCAATGGAGAGAAATATCTGGATAAGTTGAATTGGG 484

QY 53 LeuPheSer-----HisHisTrp-----Ile-----Gln---Glu 61  
Db 485 ATCTTCACACAGCCACACCATTTGGATTTTCTTCATATTTTCCATGCCATTGCCAGTCA 544

QY 62 SerLeu---Leu-----CysPro----- 66  
Db 545 ---CTAGGCTCAGTTACTTTTTCAGTTTGGCCAAATGAGCCCATCTCAACATTTGGCAG 601

QY 67 ---ProSer-----Pro-----Lys-----Glu---Val 72  
Db 602 TCCTTACCAAGCAACTACTTCACTGTATGGCCTGCAACAACCTTCTGCAATTCAGAGATC 661

QY 73 ---ThrCys-----Arg---Glu-----Met----- 77  
Db 662 CATGCTTGCTCTGGCCATGGTTAGTCTGGAACATGGCGAAACTCATTCCTGATGCGTTCT 721

QY 78 LeuThr----- 79  
Db 722 CTTACAATTGACTGGCTTCAGAAAGGCCACAGATGGATAGTCCCAAGTGAATCCATGG 781

QY 80 ---Gly-----Gly-----Cys----- 82  
Db 782 TCGGGAGCTAGAGGCACCTTCACTGCATATCTATGCGAAATATCCTGTCTTTGAAATCCG 841

QY 83 -----Leu-----Pro-----Trp-----AlaThrArg 88  
Db 842 TAATGATTTAAGGACACCCCTAAAGCAACCCGTGGGTACCTGGTACACGATGGTGTGAGA 901

QY 89 -----SerHisLeuGlyArgArgLysCys-----Ser 97  
Db 902 ATTAAGTCGACCTGTACACGGCGCAATGTAAACGGGACAAAGAGC 946

Search completed: May 11, 2003, 19:06:29  
Job time : 1062 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 16:05:46 ; Search time 360.071 Seconds  
(without alignments)  
1293.204 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delopt 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4101024

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 60 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_174132\_704/app\_query.fasta\_1.462  
-DB=GenEmbl -QFMT=fastap -SUFFIX=oligo.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=60  
-DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=45 -MODE=LOCAL -OUTFMT=ptp  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn\_1\_1758\_erunat\_05052003\_174132\_704 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rnd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Result No.	Score	Query Match	Length	DB	ID	Description
1	16	100.0	337	6	AX321911	AX321911 Sequence
2	16	100.0	2239	6	AX321909	AX321909 Sequence
3	16	100.0	5981	6	AX321910	AX321910 Sequence
4	16	100.0	161280	9	AC093903	AC093903 Homo sapi
5	8	50.0	171367	9	AL355977	AL355977 Human DNA
6	8	50.0	176623	2	AC115498	AC115498 Rattus no
7	7	43.8	579	6	AX390177	AX390177 Sequence
8	7	43.8	704	8	TOMRNAS	D17322 Tomato mRNA
9	7	43.8	720	3	AF082530	AF082530 Entamoeba
10	7	43.8	730	8	S65047	S65047 S11a-glycop
11	7	43.8	777	8	TOMRNASB	D17323 Tomato mRNA
12	7	43.8	826	6	E08516	E08516 DNA encodin
13	7	43.8	826	6	E12586	E12586 CDNA encodi
14	7	43.8	826	8	S65048	S65048 self-incomp
15	7	43.8	837	3	AF205061	AF205061 Dictyoste
16	7	43.8	1209	9	SHSDB12	AF057731 Homo sapi
17	7	43.8	1283	8	S81597	S81597 SI (S11)-se
18	7	43.8	1402	3	SPEATSBP	D38160 Sacrophaga
19	7	43.8	1487	1	FNPOPE1	X72583 Fusobacteri
20	7	43.8	1559	8	AF105149	AF105149 Zea mays
21	7	43.8	2241	6	E12585	E12585 gDNA encodi
22	7	43.8	2241	8	LPU28796	U28796 Lycopersico
23	7	43.8	3086	8	AF043091	AF043091 Hordeum v
24	7	43.8	5886	6	AX347114	AX347114 Sequence
25	7	43.8	5895	3	MITPCOCB	Z23263 Theileria p
26	7	43.8	8237	1	AWU10505	U10505 Acetobacter
27	7	43.8	8576	6	AX347131	AX347131 Sequence
28	7	43.8	9265	6	AX339183	AX339183 Sequence
29	7	43.8	9265	6	AX347055	AX347055 Sequence
30	7	43.8	11046	1	AE001096	AE001096 Archaeogl
31	7	43.8	11995	1	AE007643	AE007643 Clostridi
32	7	43.8	12971	1	AE001070	AE001070 Archaeogl
33	7	43.8	15705	1	AF198256	AF198256 Haemophil
34	7	43.8	15881	6	AX251764	AX251764 Sequence
35	7	43.8	15881	6	AX344178	AX344178 Sequence
36	7	43.8	15881	6	AX345162	AX345162 Sequence
37	7	43.8	15881	6	AX348575	AX348575 Sequence
38	7	43.8	16023	2	AC017868	AC017868 Drosophil
39	7	43.8	16633	6	AX344576	AX344576 Sequence
40	7	43.8	23933	3	U42843	U42843 Caenorhabdi
41	7	43.8	31228	3	D42841	U42841 Caenorhabdi
42	7	43.8	34544	3	AF391289	AF391289 Branchios
43	7	43.8	37977	3	L16621	L16621 Caenorhabdi
44	7	43.8	41052	9	AC002499	AC002499 Human Cos
45	7	43.8	42354	9	AC003032	AC003032 Human Chr
46	7	43.8	43309	9	AC002498	AC002498 Human Cos
47	7	43.8	44419	9	AC006506	AC006506 Homo sapi
48	7	43.8	55227	2	AC100331	AC100331 Mus muscu
49	7	43.8	58518	3	AC024202	AC024202 Caenorhab
50	7	43.8	58518	3	AC024861	AC024861 Caenorhab

C 51 7 43.8 65582 9 AL138733 Human DNA  
C 52 7 43.8 66789 2 AC103807 Homo sapi  
C 53 7 43.8 68369 9 AC093764 Homo sapi  
C 54 7 43.8 68837 2 AC091255 Homo sapi  
C 55 7 43.8 69479 9 AC008597 Homo sapi  
56 7 43.8 78436 2 AC131383 Strongylo  
57 7 43.8 79097 2 AC006897 Caenorhab  
58 7 43.8 83461 3 AE002743 Drosophila  
C 59 7 43.8 85678 2 AC111726 Rattus no  
60 7 43.8 89936 8 AP004532 Lotus jap

## ALIGNMENTS

RESULT 1  
AX321911  
LOCUS AX321911 337 bp DNA linear PAT 15-DEC-2001  
DEFINITION Sequence 442 from Patent WO0172295.  
ACCESSION AX321911  
VERSION AX321911.1 GI:17906521  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1  
AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,  
Indirias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,  
Mannion,J. and Kalos,M.D.  
TITLE Compositions and methods for the therapy and diagnosis of lung  
cancer  
JOURNAL Patent: WO 0172295-A 442 04-OCT-2001;  
CORIXA CORPORATION (US)

FEATURES  
source Location/Qualifiers  
1..337  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 103 a 60 c 93 g 81 t

Alignment Scores:  
Pred. No.: 5.69e-10 Length: 337  
Score: 16.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX321911 (1-337)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpPheTrp 16  
|||||  
Db 107 TTCCAGGCCAATTGTGGCATAGATTTTATCATATCTCGATTTTGG 154

RESULT 2  
AX321909  
LOCUS AX321909 2239 bp DNA linear PAT 15-DEC-2001  
DEFINITION Sequence 440 from Patent WO0172295.  
ACCESSION AX321909  
VERSION AX321909.1 GI:17906515  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1  
AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,  
Indirias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,  
Mannion,J. and Kalos,M.D.  
TITLE Compositions and methods for the therapy and diagnosis of lung  
cancer  
JOURNAL Patent: WO 0172295-A 440 04-OCT-2001;  
CORIXA CORPORATION (US)

FEATURES  
source Location/Qualifiers  
1..2239  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 1858 a 1029 c 1098 g 1996 t

## ALIGNMENTS

US-09-854-133-587 (1-16) x AX321909 (1-2239)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpPheTrp 16  
|||||  
Db 102 TTCCAGGCCAATTGTGGCATAGATTTTATCATATCTCGATTTTGG 149

RESULT 4  
AC093903/c  
LOCUS AC093903 161280 bp DNA linear PRI 01-MAR-2002  
DEFINITION Homo sapiens BAC clone RP11-733C7 from 4, complete sequence.  
ACCESSION AC093903 AC055827  
VERSION AC093903.3 GI:15920156  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 161280)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.



## AUTHORS

Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
99063792  
9847074

## REFERENCE

2 (bases 1 to 161280)  
Radonienko, M. and Kozlowski, A.  
The sequence of Homo sapiens BAC clone RP11-733C7  
Unpublished (2001)

## REFERENCE

3 (bases 1 to 161280)  
Waterston, R.H.

## AUTHORS

Direct Submission  
Submitted (10-SEP-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

## REFERENCE

4 (bases 1 to 161280)  
Waterston, R.H.

## AUTHORS

Direct Submission  
Submitted (04-OCT-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

## REFERENCE

5 (bases 1 to 161280)  
Waterston, R.

## AUTHORS

Direct Submission  
Submitted (01-MAR-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Oct 4, 2001 this sequence version replaced gi:15625016.

## COMMENT

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0733C07  
Drafting Center: WIBR  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-310A13. Actual start of this clone is at base position 1 of RP11-733C7; actual end is at base position 161280 of RP11-733C7.

Data from AC009792 was used to finish the clone, AC055827.

The sequence of AC055827 has been incorporated into AC093903.

## FEATURES

## source

## Location/Qualifiers

1. 161280  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"

## repeat\_region

/clone="RP11-733C7"  
/clone\_lib="RPCI-11"  
148. 304  
/rpt\_family="L1"

## repeat\_region

306. 406  
/rpt\_family=""(GGAA)n"

## repeat\_region

471. 525  
/rpt\_family="AT-rich"

## repeat\_region

530. 782  
/rpt\_family="L1"

## repeat\_region

685. 712  
/rpt\_family=""(TA)n"

## repeat\_region

1383. 1472  
/rpt\_family="GA-rich"

## repeat\_region

1697. 1836  
/rpt\_family="L1"

## repeat\_region

1837. 1956  
/rpt\_family="L1"

## repeat\_region

3005. 3221  
/rpt\_family="MIR"

## repeat\_region

4013. 4321  
/rpt\_family="Alu"

## repeat\_region

4295. 4325  
/rpt\_family=""(A)n"

## repeat\_region

5625. 5960  
/rpt\_family="L1"

## repeat\_region

5938. 5982  
/rpt\_family="A-rich"

## repeat\_region

6062. 6195  
/rpt\_family="MIR"

## repeat\_region

6388. 6462  
/rpt\_family="L2"

## repeat\_region

6546. 6868  
/rpt\_family="L1"

## misc\_feature

6866. 7318  
/note="similar to EST C06175 (NID: g1502951)"

## repeat\_region

6866. 6898  
/rpt\_family="AT-rich"

## misc\_feature

7134. 7385  
/note="similar to EST BG989897 (NID: g14393967)"

## misc\_feature

7777. 7918  
/note="similar to EST BG388107 (NID: g13281553)"

## repeat\_region

8483. 8522  
/rpt\_family="AT-rich"

## repeat\_region

8721. 9018  
/rpt\_family="Alu"

## repeat\_region

9459. 9741  
/rpt\_family="L2"

## repeat\_region

10217. 10246  
/rpt\_family="AT-rich"

## repeat\_region

10304. 10385  
/rpt\_family="MIR"

## repeat\_region

10835. 10878  
/rpt\_family="AT-rich"

## misc\_feature

11861. 11987  
/note="similar to EST BG388107 (NID: g13281553)"

## misc\_feature

11873. 11987  
/note="similar to EST BF286052 (NID: g11217122)"

## repeat\_region

12516. 12539  
/rpt\_family=""(T)n"

## repeat\_region

12929. 13232  
/rpt\_family="Alu"

## repeat\_region

13237. 13274  
/rpt\_family=""(TA)n"

## repeat\_region

13275. 13304  
/rpt\_family=""(T)n"

```
repeat_region 13696..13838
/rpt_family="MIR"
repeat_region 14127..14153
/rpt_family="AT_rich"
repeat_region 14132..14436
/rpt_family="Alu"
repeat_region 15823..16117
/rpt_family="Alu"
misc_feature 17329..17894
/note="similar to EST BG388107 (NID:G13281553)"
misc_feature 17329..17618
/note="similar to EST BF286052 (NID:G11217122)"
misc_feature 17489..17611
/note="similar to EST AI313891 (NID:G4029010)"
misc_feature 17548..17869
/note="similar to EST BF095483 (NID:G10901193)"
misc_feature 17772..18329
/note="similar to EST BG181259 (NID:G13703031)"
repeat_region 17845..17869
/rpt_family="(CTG)n"
repeat_region 18792..18828
/rpt_family="AT_rich"
repeat_region 20734..20784
/rpt_family="MERL_type"
repeat_region 20832..20944
/rpt_family="MERL_type"
repeat_region 20948..21964
/rpt_family="L1"
repeat_region 21964..21986
/rpt_family="AT_rich"
repeat_region 21965..22276
/rpt_family="Alu"
repeat_region 22277..22323
/rpt_family="L1"
repeat_region 22341..22681
/rpt_family="L1"
repeat_region 22685..22824
/rpt_family="Alu"
repeat_region 22805..22824
/rpt_family="(A)n"
repeat_region 22826..23575
/rpt_family="L1"
repeat_region 23591..23735
/rpt_family="MERL_type"
```

```
Alignment Scores:
Pred. No.: 1.4e-07 Length: 161280
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
```

US-09-854-133-587 (1-16) x AC093903 (1-161280)

```
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
|||||
Db 27653 TTCCAGGCCAATTGTCGATAGATTTTATCATATCTCGATTTTGG 27606
```

```
RESULT 5
AL355977
LOCUS AL355977 171367 bp DNA linear PRI 09-MAR-2001
DEFINITION Human DNA sequence from clone Rp11-735G18 on chromosome 10,
complete sequence.
ACCESSION AL355977
VERSION AL355977.11 GI:12956926
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 171367)
AUTHORS Lawlor,S.
```

TITLE Direct Submission  
JOURNAL Submitted (06-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
requests: clonerequest@sanger.ac.uk  
COMMENT On Feb 16, 2001 this sequence version replaced gi:12584715.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em:, EMBL; Sw:,  
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 10, constructed by the Sanger Centre Chromosome 10  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr10  
Rp11-735G18 is from the library RPCI-11.3 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACE3.6

This sequence is the entire insert of clone Rp11-735G18 The true  
left end of clone Rp11-162E8 is at 131837 in this sequence. The  
true right end of clone Rp11-344N19 is at 88698 in this sequence.

FEATURES  
source

1..171367  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="Rp11-735G18"  
/clone\_lib="RPCI-11.3"  
229..294  
/note="MIR repeat: matches 83..145 of consensus"

repeat\_region 398..636  
/note="AluJb repeat: matches 1..298 of consensus"  
repeat\_region 970..1030  
/note="MIR repeat: matches 78..142 of consensus"  
repeat\_region 1126..1233  
/note="MIR repeat: matches 22..130 of consensus"  
repeat\_region 1234..1399  
/note="FRAM repeat: matches 1..166 of consensus"  
repeat\_region 1424..1586  
/note="L2 repeat: matches 2596..2750 of consensus"  
repeat\_region 1729..2627  
/note="L2 repeat: matches 1802..2750 of consensus"  
repeat\_region 2624..2746  
/note="MIR repeat: matches 138..262 of consensus"  
repeat\_region 4276..4319  
/note="22 copies 2 mer ac 97% conserved"  
repeat\_region 4456..4604  
/note="AluY repeat: matches 33..182 of consensus"  
repeat\_region 5106..5277  
/note="MER63A repeat: matches 29..210 of consensus"  
repeat\_region 5394..5559  
/note="MER58C repeat: matches 48..87 of consensus"  
repeat\_region 5746..5823  
/note="39 copies 2 mer ta 76% conserved"  
repeat\_region 5898..8370  
/note="L1MB7 repeat: matches 3641..6171 of consensus"  
repeat\_region 9068..9347  
/note="L1PB3 repeat: matches 5854..6150 of consensus"  
repeat\_region 9582..9943  
/note="THE1B repeat: matches 1..364 of consensus"

us-09-854-133-587.oligo.rge

Tue May 13 12:12:57 2003

```

repeat_region /note="L1MB6 repeat: matches 5622. .5798 of consensus"
36838. .36986
/note="AluSg/x repeat: matches 156. .304 of consensus"
36987. .37297
/note="L1MB6 repeat: matches 5257. .5567 of consensus"
37379. .37528
/note="L1MA10 repeat: matches 6174. .6319 of consensus"
37532. .37690
/note="L1ME repeat: matches 5520. .5675 of consensus"
37686. .38146
/note="L1MB6 repeat: matches 4738. .5193 of consensus"
38147. .38436
/note="AluJo repeat: matches 1. .287 of consensus"
38437. .38768
/note="L1MB6 repeat: matches 4405. .4738 of consensus"
38821. .39104
/note="AluSc repeat: matches 1. .288 of consensus"
40351. .40570
/note="L2 repeat: matches 2538. .2750 of consensus"
40625. .40996
/note="L1PB3 repeat: matches 5758. .6145 of consensus"
41029. .41836
/note="L2 repeat: matches 1599. .2419 of consensus"
41986. .42270
/note="AluJo repeat: matches 1. .287 of consensus"
42405. .42559
/note="AluSg/x repeat: matches 147. .301 of consensus"
43254. .43541
/note="L2 repeat: matches 2387. .2686 of consensus"
43539. .43966
/note="L2 repeat: matches 1828. .2203 of consensus"
43967. .44396
/note="MSTB repeat: matches 1. .426 of consensus"
44397. .44547
/note="L2 repeat: matches 1271. .1828 of consensus"
45532. .45687
/note="L2 repeat: matches 2370. .2529 of consensus"
46297. .46712
/note="MLT2FB repeat: matches 1. .407 of consensus"
46804. .46894
/note="L2 repeat: matches 2663. .2750 of consensus"
47646. .48024
repeat_region

```

```

Alignment Scores:
Pred. NO.: 181
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 50.00%
DB: 9
Length: 171367
Matches: 8
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

```

US-09-854-133-587 (1-16) x AL355977 (1-171367)

```

QY 8 AspPhellellepHeHrrPillepHe 15
|||||
Db 36250 GATTTTATTTATTTTGGATTTT 36273

```

RESULT 6

AC115498

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC115498 176623 bp DNA linear HTG 13-JUL-2002  
Rattus norvegicus clone CH230-85C21, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 74 unordered pieces.

AC115498 GI:21736960

HTG; HTGS\_PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 176623)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

```

repeat_region 10001. .10127
/note="MIR repeat: matches 20. .147 of consensus"
10152. .10480
/note="L2 repeat: matches 2062. .2458 of consensus"
10857. .11167
/note="AluJo repeat: matches 5. .310 of consensus"
11367. .11443
/note="MIR repeat: matches 77. .154 of consensus"
12813. .13212
/note="MLT1D repeat: matches 23. .500 of consensus"
15440. .15892
/note="L2 repeat: matches 1630. .2092 of consensus"
16126. .16201
/note="L2 repeat: matches 2335. .2412 of consensus"
18939. .19041
/note="MIR repeat: matches 158. .262 of consensus"
19709. .19791
/note="MADE1 repeat: matches 1. .80 of consensus"
20433. .20476
/note="22 copies 2 mer ga 81% conserved"
20592. .20901
/note="L1MB5 repeat: matches 5578. .5887 of consensus"
20912. .21222
/note="AluJo repeat: matches 3. .312 of consensus"
21264. .21536
/note="AluSc repeat: matches 17. .292 of consensus"
21537. .21808
/note="L1MB5 repeat: matches 5896. .6166 of consensus"
23544. .23998
/note="L1MB3 repeat: matches 5737. .6176 of consensus"
24311. .24394
/note="MIR repeat: matches 43. .140 of consensus"
25430. .25854
/note="MSTA repeat: matches 1. .426 of consensus"
27916. .28009
/note="47 copies 2 mer tt 62% conserved"
28017. .28405
/note="L1P repeat: matches 4994. .5386 of consensus"
28591. .28862
/note="AluY repeat: matches 40. .307 of consensus"
28892. .29188
/note="AluSx repeat: matches 3. .297 of consensus"
29270. .30294
/note="L1PA7 repeat: matches 5131. .6141 of consensus"
30291. .30482
/note="L1PA7 repeat: matches 5055. .5130 of consensus"
30358. .30478
/note="L1MC1 repeat: matches 6206. .6325 of consensus"
30479. .31518
/note="L1MC1 repeat: matches 4921. .6092 of consensus"
31519. .31811
/note="AluY repeat: matches 1. .295 of consensus"
31812. .32673
/note="L1MC1 repeat: matches 4078. .4921 of consensus"
32696. .33272
/note="MER72 repeat: matches 82. .642 of consensus"
33274. .33876
/note="L1MC1 repeat: matches 3473. .4089 of consensus"
33893. .34122
/note="MER2 repeat: matches 114. .344 of consensus"
34124. .34429
/note="L1MC1 repeat: matches 2966. .3286 of consensus"
34627. .34911
/note="AluSx repeat: matches 13. .298 of consensus"
35007. .35300
/note="MER2 repeat: matches 11. .334 of consensus"
35897. .36199
/note="AluSx repeat: matches 1. .298 of consensus"
36262. .36423
/note="L1MB6 repeat: matches 5798. .5939 of consensus"
36424. .36617
/note="L1PA13 repeat: matches 5960. .6151 of consensus"
36618. .36787
repeat_region

```



Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,  
 Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,  
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,  
 Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.

# TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

## REFERENCE AUTHORS TITLE JOURNAL

### COMMENT

2 (bases 1 to 176623)

Worley, K.C.

Direct Submission

Submitted (20-MAR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 176623)

Worley, K.C.

Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

On Jul 12, 2002 this sequence version replaced gi:19698616.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Project Information

Center project name: GNWZ

Center clone name: CH230-85C21

Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 107353 bases at least Q40

Consensus quality: 114147 bases at least Q30

Consensus quality: 119101 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 74 contigs. The true order of the pieces

\* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 1126: contig of 1125 bp in length  
 1225: gap of unknown length  
 1226: contig of 1616 bp in length  
 2841: gap of unknown length  
 2941: contig of 1618 bp in length  
 4559: gap of unknown length  
 4560: contig of 1300 bp in length  
 5959: gap of unknown length  
 6059: contig of 1052 bp in length  
 7111: gap of unknown length  
 7211: contig of 1290 bp in length  
 8501: gap of unknown length  
 8601: contig of 1032 bp in length  
 9633: gap of unknown length  
 9733: contig of 1360 bp in length  
 11093: gap of unknown length  
 11193: contig of 1512 bp in length  
 12705: gap of unknown length  
 12805: contig of 1348 bp in length  
 14153: gap of unknown length  
 14253: contig of 1094 bp in length  
 15347: gap of unknown length  
 15447: contig of 1230 bp in length  
 16677: gap of unknown length  
 16777: contig of 1473 bp in length  
 18250: gap of unknown length  
 18350: contig of 1399 bp in length  
 19749: gap of unknown length  
 19849: contig of 1632 bp in length  
 21481: gap of unknown length  
 21581: contig of 1156 bp in length  
 22737: gap of unknown length  
 22837: contig of 1569 bp in length  
 24406: gap of unknown length  
 24506: contig of 1057 bp in length  
 25563: gap of unknown length  
 25663: contig of 1166 bp in length  
 26829: gap of unknown length  
 26929: contig of 1174 bp in length  
 28103: gap of unknown length  
 28203: contig of 1560 bp in length  
 29763: gap of unknown length  
 29863: contig of 1163 bp in length  
 31026: gap of unknown length  
 31126: contig of 1540 bp in length  
 32666: gap of unknown length  
 32766: contig of 1803 bp in length  
 34569: gap of unknown length  
 34669: contig of 1409 bp in length  
 36078: gap of unknown length  
 36178: contig of 1225 bp in length  
 37403: gap of unknown length  
 37503: contig of 1188 bp in length  
 38691: gap of unknown length  
 38791: contig of 2312 bp in length  
 41103: gap of unknown length  
 41203: contig of 1715 bp in length  
 42918: gap of unknown length  
 43018: contig of 1261 bp in length  
 44279: gap of unknown length  
 44379: contig of 1514 bp in length  
 45893: gap of unknown length  
 45993: contig of 1242 bp in length  
 47235: gap of unknown length  
 47335: contig of 1523 bp in length  
 48858: gap of unknown length  
 48958: contig of 1970 bp in length  
 50928: gap of unknown length



FEATURES source Location/Qualifiers  
1..579  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 195 a 123 c 100 g 161 t  
ORIGIN  
Alignment Scores: 15.5 Length: 579  
Pred. No.: 7.00 Matches: 7  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 43.75% Gaps: 0  
DB:  
US-09-854-133-587 (1-16) x AX390177 (1-579)  
QY 10 IlellePheTrpIlePheTrp 16  
|||||  
Db 328 ATATATTTTGGATATTTTGG 308  
RESULT 8  
TOMRNASA/c 704 bp mRNA linear PLN 04-FEB-1999  
LOCUS  
DEFINITION Tomato mRNA for Slla-RNase, partial sequence.  
ACCESSION D17322  
VERSION D17322.1 GI:443776  
KEYWORDS RNase; ribonuclease; style tissue specific Slla-RNase.  
SOURCE Lycopersicon peruvianum  
ORGANISM Lycopersicon peruvianum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
REFERENCE 1 (bases 1 to 704)  
AUTHORS Chung, I., Nakata, K., Tanaka, H., Ito, T., Horiuchi, H., Ohta, A. and Takagi, M.  
TITLE Identification of cDNA clones coding for the style specific Slla-RNase gene associated with gametophytic self-incompatibility in tomato (Lycopersicon peruvianum)  
JOURNAL Biosci. Biotechnol. Biochem. (1993) In press  
REFERENCE 2 (sites)  
AUTHORS Chung, I.K., Ito, T., Tanaka, H., Ohta, A., Nan, H.G. and Takagi, M.  
TITLE Molecular diversity of three S-allele cDNAs associated with gametophytic self-incompatibility in Lycopersicon peruvianum  
JOURNAL Plant Mol. Biol. 25 (2), 757-762 (1994)  
MEDLINE 95036055  
PUBMED 7948929  
REFERENCE 3 (bases 1 to 704)  
AUTHORS Chung, I.-K.  
TITLE Direct Submission  
JOURNAL Submitted (28-JUL-1993) Il-Kyung Chung, The University of Tokyo, Dept. of Agricultural Chemistry, Bunkyo-ku, Tokyo 113, Japan (Tel: 03-3812-2111 (ex.3085), Fax: 03-3812-9246)  
COMMENT On Jan 22, 1994 this sequence version replaced gi:391924. Submitted (28-JUL-1993) to DDBJ by: Il-Kyung Chung  
Dept. of Agricultural Chemistry  
The University of Tokyo  
Bunkyo-ku,  
Tokyo 113  
Japan  
Phone: 03-3812-2111 x3085  
Fax: 03-3812-9246.  
FEATURES source Location/Qualifiers  
1..704  
/organism="Lycopersicon peruvianum"  
/db\_xref="taxon:4082"  
/tissue\_type="mature style"  
<1..489  
/codon\_start=1  
/product="Slla-RNase"  
/protein\_id="BAA04144.1"

50929 51028: gap of unknown length  
51029 52476: contig of 1448 bp in length  
52477 52576: gap of unknown length  
52577 55444: contig of 2868 bp in length  
55445 55445: gap of unknown length  
55445 56856: contig of 1312 bp in length  
56857 56956: gap of unknown length  
56957 58542: contig of 1586 bp in length  
58543 58642: gap of unknown length  
58643 60626: contig of 1984 bp in length  
60627 60726: gap of unknown length  
60727 62590: contig of 1864 bp in length  
62591 62690: gap of unknown length  
62691 65084: contig of 2394 bp in length  
65085 65184: gap of unknown length  
65185 66974: contig of 1790 bp in length  
66975 67074: gap of unknown length  
67075 69013: contig of 1939 bp in length  
69014 69113: gap of unknown length  
69114 70953: contig of 1840 bp in length  
70954 71053: gap of unknown length  
71054 73696: contig of 2643 bp in length  
73697 73796: gap of unknown length  
73797 76005: contig of 2209 bp in length  
76006 76105: gap of unknown length  
76106 78643: contig of 2538 bp in length  
78644 78743: gap of unknown length  
78744 82025: contig of 3282 bp in length  
82026 82125: gap of unknown length  
82126 83983: contig of 1858 bp in length  
83984 84083: gap of unknown length  
84084 86082: contig of 1999 bp in length  
86083 86182: gap of unknown length  
86183 88378: contig of 2196 bp in length  
88379 88478: gap of unknown length  
88479 91276: contig of 2798 bp in length  
91277 91376: gap of unknown length  
91377 94059: contig of 2683 bp in length  
94060 94159: gap of unknown length  
94160 96805: contig of 2646 bp in length

Alignment Scores:  
Pred. No.: 185 Length: 176623  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 50.00% Indels: 0  
DB: 2 Gaps: 0

US-09-854-133-587 (1-16) x AC115498 (1-176623)

QY 5 CysGlyIleAspPheIlePhe 12  
|||||  
Db 128732 TGTCATAGATTTCATCATTTTC 128755

RESULT 7  
AX390177/c AX390177 579 bp DNA linear PAT 19-MAR-2002  
LOCUS  
DEFINITION Sequence 5105 from Patent WO0214500.  
ACCESSION AX390177  
VERSION AX390177.1 GI:19583307  
KEYWORDS  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1

AUTHORS Escobedo, J., Garcia, P.D., Sudduth-Klinger, J., Reinhard, C., Pot, D. and Randazzo, F., Lamson, G., Scott, E.M., Zhang, G., Kassam, A., Pot, D. and Labat, I.  
TITLE Human genes and gene expression products  
JOURNAL Patent: WO 0214500-A 5105 21-FEB-2002;  
CHIRON CORPORATION (US) ; Hyseq Inc. (US)

/db\_xref="GI:443777"  
/translation="HGLWPDIKGTILNNCNPDAKYASVTGGKFKVRNKHWPDLILTEA  
ASLNSQGFWAYQFKKHGTCSDLENFQEKYFDLALILKDKFDLLTFRNKGIIIPKSTCT  
INKIQKTIRTVTGVVPLNSCTPTMELLEVGICFNRDASKLIDCDQPKTCDTSGNTEIF  
FG"  
polyA\_site 235 a 136 c 124 g 209 t  
BASE COUNT 704  
ORIGIN

Alignment Scores:  
Pred. No.: 18.4 Length: 704  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 8 Gaps: 0

US-09-854-133-587 (1-16) x TOMRNASA (1-704)

QY 6 GlyIleAspPheIleIlePhe 12

Db 550 GGAATTGATTATTATATTATTT 530

RESULT 9  
AF082530

LOCUS AF082530

DEFINITION Entamoeba histolytica Grainin 2 (grainin 2) mRNA linear INV 12-DEC-2000

ACCESSION AF082530

VERSION AF082530.2 GI:7266947

KEYWORDS

SOURCE Entamoeba histolytica.

ORGANISM

REFERENCE Entamoeba histolytica.

AUTHORS Eukaryota; Entamoebidae; Entamoeba.

TITLE Nickel, R., Jacobs, T., Urban, B., Scholze, H., Bruhn, H. and Leippe, M.

JOURNAL Two novel calcium-binding proteins from cytoplasmic granules of the  
MEDLINE protozoan parasite Entamoeba histolytica  
PUBMED FEBS Lett. 486 (2), 112-116 (2000)

REFERENCE 20565533

AUTHORS 11113449

TITLE Nickel, R. and Leippe, M.

JOURNAL Direct Submission

Submitted (05-AUG-1998) Molecular Biology, Bernhard Nocht Institute  
for Tropical Medicine, Bernhard-Nocht-Str. 74, Hamburg 20359,  
Germany

Sequence update by submitter

On Mar 20, 2000 this sequence version replaced gi:6523790.

Location/Qualifiers

1. .720

/organism="Entamoeba histolytica"

/strain="HM-1:IMSS"

/db\_xref="taxon:5759"

1. .720

/gene="grainin 2"

10. .651

/gene="grainin 2"

/codon\_start=1

/product="Grainin 2"

/protein\_id="AAF14855.2"

/db\_xref="GI:7266948"

/translation="MSLFAIQAAADAFVQMIQAANSDPNLKFQWFLVERLDAKD  
LQNLQSWFISVDKSGTLEIGELKKAFCGKIKVDDKTKRLMRVFDIDMSGSIGFF  
EFLALWNFMNLCNETFKHFDADKSGSLDVNELIKALPMLGFCNKRSDVLLKMGSS  
LGSKVKVKNQFISTAAAYLGQCRSIYOKTFNMKREIDNAEFDKFVNLVLALSG"

BASE COUNT 264 a 102 c 124 g 230 t

ORIGIN

# Alignment Scores:

Pred. No.: 18.8 Length: 720  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 3 Gaps: 0

US-09-854-133-587 (1-16) x AF082530 (1-720)

QY 1 PheGlnAlaAsnCysGlyIle 7

Db 641 TTTCAGGCTAATTGCGGTATT 661

RESULT 10

S65047/c

LOCUS S65047

DEFINITION S65047

S65047  
S65047.1 GI:410000  
Lycopersicon peruvianum style PI 126441 Slla-plant.  
{clone I} Lycopersicon peruvianum-tomatoes, PI 126441, style,  
Slla-plant, mRNA Partial, 730 nt].

ACCESSION S65047

VERSION S65047

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REMARK

GenBank staff at the National Library of Medicine created this  
entry [NCBI gibbsq 136759] from the original journal article.  
This sequence comes from Fig. 1.

Location/Qualifiers

1. .730

/organism="Lycopersicon peruvianum"

/db\_xref="taxon:4082"

1. .730

/partial

/gene="Slla-glycoprotein"

1. .489

/partial

/gene="Slla-glycoprotein"

/note="self-incompatibility associated S-glycoprotein;"

S-associated RNase; This sequence comes from Fig. 1"

/codon\_start=1

/product="Slla-glycoprotein"

/protein\_id="AAC60562.1"

/db\_xref="GI:410001"

/translation="HGLWPDIKGTILNNCNPDAKYASVTGGKFKVRNKHWPDLILTEA  
ASLNSQGFWAYQFKKHGTCSDLENFQEKYFDLALILKDKFDLLTFRNKGIIIPKSTCT  
INKIQKTIRTVTGVVPLNSCTPTMELLEVGICFNRDASKLIDCDQPKTCDTSGNTEIF  
FP"

BASE COUNT 261 a 138 c 122 g 209 t

ORIGIN

# Alignment Scores:

Pred. No.: 19 Length: 730  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 8 Gaps: 0

us-09-854-133-587.oligo.rge

Tue May 13 12:12:57 2003

US-09-854-133-587 (1-16) x S65047 (1-730)

QY 6 GlyIleAspPheIleIlePhe 12  
 Db 550 GGAATTGATTTTATATATT 530

RESULT 11  
 TOMRNASB/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

777 bp mRNA linear PLN 04-FEB-1999  
 Tomato mRNA for S11a-RNase, partial sequence.  
 D17323  
 D17323.1 GI:443778  
 RNase; ribonuclease; style tissue specific S11a-RNase.  
 Lycopersicon peruvianum mature style cDNA to mRNA, clone II.  
 Lycopersicon peruvianum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

1 (bases 1 to 777)  
 Chung, I., Nakata, K., Tanaka, H., Ito, T., Horiuchi, H., Ohta, A. and Takagi, M.  
 Identification of cDNA clones coding for the style specific S11a-RNase gene associated with gametophytic self-incompatibility in tomato (Lycopersicon peruvianum)  
 Biosci. Biotechnol. Biochem. (1993) In press  
 2 (bases 1 to 777)  
 Chung, I.-K.  
 Direct Submission  
 Submitted (28-JUL-1993) IL-Kyung Chung, The University of Tokyo, Dept. of Agricultural Chemistry, Bunkyo-ku, Tokyo 113, Japan  
 (Tel: 03-3812-2111 (ex. 3085), Fax: 03-3812-9246)  
 On Jan 22, 1994 this sequence version replaced gi:391925.  
 Submitted (28-JUL-1993) to DDBJ by:  
 IL-Kyung Chung  
 Dept. of Agricultural Chemistry  
 The University of Tokyo  
 Bunkyo-ku,  
 Tokyo 113  
 Japan  
 Phone: 03-3812-2111 x3085  
 Fax: 03-3812-9246.

FEATURES  
 source  
 CDS  
 polyA\_site  
 BASE COUNT  
 ORIGIN

1..777  
 /organism="Lycopersicon peruvianum"  
 /db\_xref="taxon:4082"  
 /tissue\_type="mature style"  
 <1..489  
 /codon\_start=1  
 /product="S11a-RNase"  
 /protein\_id="BAA04145.1"  
 /db\_xref="GI:443779"  
 /translation="HGLWPDIKGTILNCPDPAKYASVTGGKFKRNKHWPDLILTEA  
 ASLNSQGFWAYQFKKHGTCCSDFNQEKYFDLALILKDFDLTFRNKGIIIPKSTCT  
 INKIQKTIRTVTGVPVNLSCPTMELLEVCICFNDRASKLIDCDQPKTCDTSGNTEIF  
 FG"  
 777  
 260 a 150 c 131 g 236 t

Alignment Scores:  
 Pred. No.: 20.1 Length: 777  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 8 Gaps: 0

US-09-854-133-587 (1-16) x TOMRNASB (1-777)  
 QY 6 GlyIleAspPheIleIlePhe 12  
 Db 550 GGAATTGATTTTATATATT 530

RESULT 12

E08516/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

826 bp RNA linear PAT 29-SEP-1997  
 DNA encoding style-specific S-ribonuclease..  
 E08516  
 E08516.1 GI:2176631  
 JP 1994335389-A/1.  
 Lycopersicon peruvianum.  
 Lycopersicon peruvianum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

1 (bases 1 to 826)  
 Tei, I., Nakada, K., Ito, T., Horiuchi, H., Ota, A., Takagi, M., Tsubura, H., Tanaka, H. and Ishiguro, Y.  
 S-RIBONUCLEASE SPECIFIC TO STYLE AND DNA SEQUENCE CODING THEREFOR  
 Patent: JP 1994335389-A 1 06-DEC-1994;  
 KAGOME CO LTD  
 OS Lycopersicon peruvianum (tomato)  
 PN JP 1994335389-A/1  
 PD 06-DEC-1994  
 PF 27-MAY-1993 JP 1993126286  
 PI TEI ITSUIRU, NAKADA KENGO, ITO TORU, HORIUCHI HIROYUKI, PI  
 OTA AKINORI,  
 PI TAKAGI MASAMICHI, TSUBURA HIROKAZU, TANAKA HIROSHI, PI  
 ISHIGURO YUKIO  
 PC C12N9/22,C12N15/52;  
 CC strandedness: Double;  
 CC topology: Linear;  
 FH Key Location/Qualifiers  
 FH source 1..826  
 /organism="Lycopersicon peruvianum" FT  
 /tissue\_type="pistil"  
 /mat\_peptide 1..486  
 /product="style-specific S-ribonuclease".

FEATURES  
 source  
 BASE COUNT  
 ORIGIN

1..826  
 /organism="Lycopersicon peruvianum"  
 /db\_xref="taxon:4082"  
 309 a 150 c 131 g 236 t

Alignment Scores:  
 Pred. No.: 21.2 Length: 826  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x E08516 (1-826)  
 QY 6 GlyIleAspPheIleIlePhe 12  
 Db 550 GGAATTGATTTTATATATT 530

RESULT 13  
 E12586/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

E12586  
 CDNA encoding S-ribonuclease.  
 E12586  
 E12586.1 GI:3251418  
 JP 1997028381-A/3.  
 Lycopersicon peruvianum.  
 Lycopersicon peruvianum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.  
 1 (bases 1 to 826)

ence comes from Fig  
Location/Qualifiers  
1. .826

**Lignment Scores:**

Alignment Scores:



Pred. No.: 21.5 Length: 837  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 3 Gaps: 0

US-09-854-133-587 (1-16) x AF205061 (1-837)

QY 9 PheillePheTrpIlePhe 15  
|||||  
Db 471 TTTATAATTTTGGATCTTT 451

RESULT 16  
HSHSDB12 1209 bp DNA linear PRI 20-JAN-1999  
LOCUS Homo sapiens 17-beta-hydroxysteroid dehydrogenase IV (HSD17B4)  
DEFINITION gene, exon 13.  
ACCESSION AF057731  
VERSION AF057731.1 GI:4165038  
KEYWORDS  
SEGMENT  
SOURCE  
ORGANISM Homo sapiens.  
Homo sapiens.  
Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1209)  
Leenders,F., Dolez,V., Begue,A., Moller,G., Gloeckner,J.C., de  
Launoit,Y. and Adamski,J.  
Structure of the gene for the human 17beta-hydroxysteroid  
dehydrogenase type IV  
Mamm. Genome 9 (12), 1036-1041 (1998)  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE 9909251  
9880674  
2 (bases 1 to 1209)  
Leenders,F., Dolez,V., Begue,A., Moeller,G., Gloeckner,J.C., de  
Launoit,Y. and Adamski,J.  
Direct Submission  
Submitted (07-APR-1998) Institute of Mammalian Genetics, GSF -  
Research Centre for Health and Environment, Ingolstaedter Landstr.  
1, Neuherberg 85764, Germany  
Location/Qualifiers  
1..1209  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/map="5q21"  
526..762  
/gene="HSD17B4"  
/number=13  
BASE COUNT 402 a 194 c 215 g 398 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 29.8 Length: 1209  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x HSHSDB12 (1-1209)

QY 9 PheillePheTrpIlePhe 15  
|||||  
Db 217 TTTATAATTTTGGATCTTT 237

RESULT 17  
S81597/c  
LOCUS  
DEFINITION SI (S11)=self-incompatibility/prevents self-fertilization  
[Lycopersicon peruvianum, Genomic, 1283 nt].  
ACCESSION S81597

VERSION S81597.1 GI:1478372  
KEYWORDS  
SOURCE Lycopersicon peruvianum.  
ORGANISM Lycopersicon peruvianum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 1283)  
Chung,I.K., Lee,S.Y., Ito,T., Tanaka,H., Nam,H.G. and Takagi,M.  
The 5' flanking sequences of two S alleles in Lycopersicon  
peruvianum are highly heterologous but contain short blocks of  
homologous sequences  
Plant Cell Physiol. 36 (8), 1621-1627 (1995)  
JOURNAL  
MEDLINE  
PUBMED  
REMARK GenBank staff at the National Library of Medicine created this  
entry [NCBI gibbsq 176855] from the original journal article.  
This sequence comes from Fig. 2.

FEATURES  
Location/Qualifiers  
1..1283  
/organism="Lycopersicon peruvianum"  
/db\_xref="taxon:4082"  
448..1185  
/gene="SI"  
/note="self-incompatibility/prevents self-fertilization"  
/allele="S11"  
join(448..678,772..1185)  
/gene="SI"  
/note="RNase; This sequence comes from Fig. 2"  
/codon\_start=1  
/protein\_id="AAB36131.1"  
/db\_xref="GI:1478373"  
/translation="MFKTQHTLAFFILLCALPDVYGFNQLQLVLRWPASECKGKKE  
RTPNFTIHGLWPDIKGTILNKNPDAKYASVTGGKFKRNKHPDLILTEAASLNSQ  
GFWAYQFKKHGTCSDLFNQEKYFDLALILKDFDLTTFRNKGIIIPKSTCTINKIQK  
TIRTVGVVPLNSCTPTMELLEVGICFNDRDASKLIDCDQPKTCDTSGNTEIFFP"

BASE COUNT 427 a 234 c 204 g 418 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 31.4 Length: 1283  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 8 Gaps: 0

US-09-854-133-587 (1-16) x S81597 (1-1283)

QY 6 GlyIleAspPheIleIlePhe 12  
|||||  
Db 1245 GGAATTGATTTTATTATTT 1225

RESULT 18  
SPEATSBP/c  
LOCUS  
DEFINITION SPEATSBP  
Sacrophaga peregrina ATBP mRNA for A/T-stretch binding protein  
(transcription factor), complete cds.  
ACCESSION D38160  
VERSION D38160.1 GI:1754528  
KEYWORDS A/T-stretch binding protein; transcription factor; ATBP.  
SOURCE Sacrophaga peregrina  
ORGANISM Sacrophaga peregrina  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Oestroidea; Sarcophagidae; Sarcophaga.  
1 (bases 1 to 1402)  
Nakanishi-Matsui,M., Kubo,T. and Natori,S.  
Molecular cloning and nuclear localization of ATBP, a novel  
(A+T)-stretch-binding protein of Sarcophaga peregrina (flesh fly)  
Eur. J. Biochem. 230 (2), 396-400 (1995)  
JOURNAL  
MEDLINE 95331271

REFERENCE 2 (bases 1 to 1402)  
AUTHORS Nakanishi-Matsui, M.  
TITLE Direct Submission  
JOURNAL Submitted (03-SEP-1994) Mayumi Nakanishi-Matsui, Faculty of  
Pharmaceutical Sciences, University of Tokyo; 7-3-1, Hongo,  
Bunkyo-ku, Tokyo 113, Japan (Tel:03-3812-2111(ex.4820),  
Fax:03-5684-2973)  
COMMENT On Dec 27, 1996 this sequence version replaced gi:1065611.  
FEATURES Location/Qualifiers  
source 1..1402  
/organism="Sarcophaga peregrina"  
/db\_xref="taxon:7386"  
186..1250  
/codon\_start=1  
/product="A/T-stretch binding protein"  
/protein\_id="BAA07349.1"  
/db\_xref="GI:1065612"  
/translation="MGFPRIISKNNKIYTKLGAFCLSGDDGQFQWIVCHTCORELQTD  
KFWKHQDEHNFMHGPKEQGRTAQAQYMEAAEAAMTPLPLRYKVSNDQQRDDVVS  
TEDEDMQKEPKDYTEMRAHDDQQQTAAVAIDIKLEPSSLSQSSAVQAQQQQQQQQQ  
QQ  
MSVAAAAAASNVQVSTMANLLPQELQYKQELQYKQELQYKQELQYKQELQYKQ  
SSDDGERFYICDFENCGLKFKYHSRLHRSVHSKVRFFACEICGASFQSCNLSLTHR  
KKKHALKGTATKATLVPSQSF"  
1402  
polyA\_site 507 a 310 c 278 g 307 t  
BASE COUNT 507 a 310 c 278 g 307 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 34 Length: 1402  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 3 Gaps: 0  
US-09-854-133-587 (1-16) x SPEATSBP (1-1402)  
QY 6 GlyIleAspPheIleIlePhe 12  
|||||  
Db 230 GGTATAGATTATTATTTT 210  
RESULT 19  
FNPOMP1/C  
LOCUS FNPOMP1 1487 bp DNA linear BCT 16-AUG-2000  
DEFINITION Fusobacterium nucleatum foma gene, strain ATCC 10953.  
ACCESSION X72583  
VERSION X72583.2 GI:9844097  
KEYWORDS 40 kDa protein; cell surface protein; foma gene; outer membrane  
protein; transmembrane protein.  
SOURCE Fusobacterium nucleatum subsp. polymorphum.  
ORGANISM Fusobacterium nucleatum subsp. polymorphum  
Bacteria; Fusobacteria; Fusobacterium.  
1 (bases 1 to 1487)  
REFERENCE Bolstad, A.I., Tommassen, J. and Jensen, H.B.  
AUTHORS Sequence variability of the 40-kDa outer membrane proteins of  
TITLE Fusobacterium nucleatum strains and a model for the topology of the  
JOURNAL proteins  
MEDLINE Mol. Gen. Genet. 244 (1), 104-110 (1994)  
PUBMED 94316187  
8041356  
REFERENCE 2 (bases 1 to 1487)  
AUTHORS Bolstad, A.I.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAR-1993) A. I. Bolstad, University of Bergen, Dept  
of Biochemistry, Arstadveien 19, 5009 Bergen, NORWAY  
REMARK revised by [3] MAT  
REFERENCE 3 (bases 1 to 1487)  
AUTHORS Bolstad, A.I.  
TITLE Direct Submission  
JOURNAL Submitted (09-AUG-1994) A. I. Bolstad, University of Bergen, Dept  
of Biochemistry, Arstadveien 19, 5009 Bergen, NORWAY  
REMARK revised by [4]

REFERENCE 4 (bases 1 to 1487)  
AUTHORS Bolstad, A.I.  
TITLE Direct Submission  
JOURNAL Submitted (29-SEP-1994) A. I. Bolstad, University of Bergen, Dept  
of Biochemistry, Arstadveien 19, 5009 Bergen, NORWAY  
REMARK revised by [5]  
REFERENCE 5 (bases 1 to 1487)  
AUTHORS Jensen, H.B.  
TITLE Direct Submission  
JOURNAL Submitted (14-AUG-2000) H.B. Jensen, University of Bergen, Dept of  
Molecular Biology, Thormohlensgate 55, 5020 Bergen, NORWAY  
COMMENT On Aug 17, 2000 this sequence version replaced gi:551439.  
FEATURES Location/Qualifiers  
source 1..1487  
/organism="Fusobacterium nucleatum subsp. polymorphum"  
/strain="ATCC 10953"  
/sub\_species="polymorphum"  
/db\_xref="taxon:76857"  
322..1447  
/gene="foma"  
322..326  
/gene="foma"  
335..1447  
/gene="foma"  
/codon\_start=1  
/translation="MKKLALVLGLLLVVGSAKEVMPAPTPAPEKVVVEYVEKPVIV  
YRDREVAWAPWPNQSVQYRWYGEVEKKNPDKDKNWATGKVNAGRIQLTKVNFT  
EKQTLVTRNRHHTLNDTDANNKSKNGAADYRLRHFYNGFKLGSSKVNATSRVEFKQ  
KTNDGEKSLGASVLFDEADYIYNNFFKVDKGLRPGYKYVWKGNGEGTPTVHNE  
YHLAFESDFTLPFNFALELDLSYNNRYREKFTTDLGLKKAETGELTAVLSNYTPLY  
KAGAFELGFNAEGGYDYNHMQYKRIGGEDGTSVDRRDYELYLEPTLQVSYKPTDFVK  
LYAAGADYRNRRITGESEVKKRWQPTASAGMKVTF"  
335..394  
/gene="foma"  
395..1444  
/gene="foma"  
/product="porin"  
/note="major outer membrane protein, 40kDa"  
628 a 181 c 254 g 424 t  
BASE COUNT 628 a 181 c 254 g 424 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 35.9 Length: 1487  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 1 Gaps: 0  
US-09-854-133-587 (1-16) x FNPOMP1 (1-1487)  
QY 9 PheIleIlePheTrpIlePhe 15  
|||||  
Db 571 TTTATCATCTTTTGGATTTT 551  
RESULT 20  
AF105149/c  
LOCUS AF105149 1559 bp mRNA linear  
DEFINITION Zea mays kaurene synthase (KS) mRNA, partial cds.  
ACCESSION AF105149  
VERSION AF105149.1 GI:4959321  
KEYWORDS Zea mays.  
SOURCE Zea mays.  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 1559)

AUTHORS Richman,A.S. and Brandle,J.E.  
TITLE Partial kaurene synthase gene from Zea mays  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1559)  
AUTHORS Richman,A.S. and Brandle,J.E.  
TITLE Direct Submission  
JOURNAL Submitted (09-NOV-1998) Southern Crop Protection and Food Research  
Centre, Agriculture and AgriFood Canada, 1391 Sandford Street,  
London, ON N5V 4T3, Canada  
FEATURES Location/Qualifiers  
source  
1..1559  
/organism="Zea mays"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="csul86"  
/tissue\_type="illuminated leaves; sheaths"  
/clone\_lib="UMC Clone Distribution Centre"  
/dev\_stage="5 weeks old"  
<1..1559  
/gene="KS"  
<1..1377  
/gene="KS"  
/function="synthesis of (-)-kaurene by ionization  
dependant cyclization of copalyl diphosphate"  
/note="terpene cyclase"  
/codon\_start=1  
/product="kaurene synthase"  
/protein\_id="AAD34319.1"  
/db\_xref="GI:4959322"  
/translation="GYNVSSDELYHVVEASGLHNSLGGYLNDRTPLELHKASTVSIS  
EDESILDSIGSRSTLLRQESGALRKPSPKFVEHALDGPYTTLDRLHHRWNIE  
NFNIEQHMLETPYLSNQHTSRDILALSIRDFSSQFTYQQLHLESWVKECRLDQL  
QFAROKLAYFLSAAGTMFSPELSDARTLWAKNGVLTITIVDDFDVAGSKELENLVM  
LVEMWDEHHKVEFYSEQVEIIFSSYDSVNLGKASLVQDRSITKHLVEIWLDLKS  
MMTEVEWRLSKYVPTKEKEYMINASLTFLGPIVLPALYFVGPKISEIVKDPYDELF  
KLMSTCGRLLNDVQTFEREYNEGLNSVLLVLHGGMSISDAKRKLQKPIDTCRRDL  
LSLVLRESVVPKCKELFWKMKVCYFFYSTDTGFSQVERAKEVDVINEPLKLGQ  
SHTLVSDV"  
BASE COUNT 453 a 288 c 362 g 456 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 37.4 Length: 1559  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 8 Gaps: 0  
US-09-854-133-587 (1-16) x AF105149 (1-1559)  
QY 9 PheilleilepHeTrpIlePhe 15  
|||||  
Db 996 TTATCATATATTCTGGATCTTT 976  
RESULT 21  
E12585/c  
LOCUS E12585  
DEFINITION gDNA encoding.  
ACCESSION E12585  
VERSION E12585.1 GI:3251417  
KEYWORDS JP 1997028381-A/2.  
SOURCE Lycopersicon peruvianum.  
ORGANISM Lycopersicon peruvianum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
REFERENCE 1 (bases 1 to 2241)  
AUTHORS Tei,I., Minami,K. and Takagi,M.  
TITLE S- RIBONUCLEASE GENE AND PROMOTER SEQUENCE  
JOURNAL Patent: JP 1997028381-A 2 04-FEB-1997;  
TEI ITSUKIYON, MINAMI KOUKICHI, TAKAGI MASAMICHI

COMMENT OS Lycopersicon peruvianum  
PN JP 1997028381-A/2  
PD 04-FEB-1997  
PF 24-JUL-1995 JP 1995187557  
PI TEI ITSUKIYON, MINAMI KOUKICHI, TAKAGI MASAMICHI PC  
C12N15/09,C07H21/04,C12N1/21//A01H1/00,C12N5/10,C12N9/22, PC  
(C12N1/21,  
PC C12R1:19);  
CC strandedness: Double;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No; Location/Qualifiers  
FH Key  
FT source 1..2241  
FT /organism="Lycopersicon peruvianum" FT  
TATA\_signal 1255..1258  
FT 5'UTR 1..1365  
FT exon 1..1596  
FT /number=1  
FT intron 1597..1687  
FT /number=1  
FT exon 1688..2241  
FT /number=2  
FT CDS join(1366..1596,1688..2101)  
FT /product="S-ribonuclease"  
FT 3'UTR 2102..2241.  
FEATURES Location/Qualifiers  
source 1..2241  
/organism="Lycopersicon peruvianum"  
/db\_xref="taxon:4082" 724 t  
BASE COUNT 683 a 390 c 444 g  
ORIGIN  
Alignment Scores:  
Pred. No.: 51.7 Length: 2241  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 6 Gaps: 0  
US-09-854-133-587 (1-16) x E12585 (1-2241)  
QY 6 GlylleAspPheIleIlePhe 12  
|||||  
Db 2203 GGAATTGATTTTATATATTT 2183  
RESULT 22  
LPU28796/c  
LOCUS LPU28796  
DEFINITION Lycopersicon peruvianum self incompatibility (sl2) gene, complete  
cds.  
ACCESSION U28796  
VERSION U28796.1 GI:1002595  
KEYWORDS  
SOURCE Lycopersicon peruvianum.  
ORGANISM Lycopersicon peruvianum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
REFERENCE 1 (bases 1066 to 2241)  
AUTHORS Chung,I.-K., Leel,S.Y., Masamichi,T. and Nam,H.G.  
TITLE 5' Flanking sequences of two S alleles in Lycopersicon peruvianum  
are highly heterologous but contain short blocks of homologous  
sequences  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2241)  
AUTHORS Chung,I.-K., Leel,S.Y., Masamichi,T. and Nam,H.G.  
TITLE Direct Submission  
JOURNAL Submitted (09-JUN-1995) Il-Kyung Chung, Life Science, Pohang  
University of Science and Technology, Jigok, Pohang, Kyungbuk,

```
FEATURES
source      Location/Qualifiers
1. .2241
/organism="Lycopersicon peruvianum"
/db_xref="taxon:4082"
join(1366. .1596,1688. .2101)
/gene="sl2"
CDS
join(1366. .1596,1688. .2101)
/gene="sl2"
/function="self incompatibility"
/codon_start=1
/product="ribonuclease"
/protein_id="AAA77040.1"
/db_xref="GI:1002596"
/translation="MFKTQHTLAFFILLCALPDVYGTFFNQLQLVLRWPAFCKGKKCE
RTPNFTIHLWPDIKGTILNCPNDAKVASVTGGKFKVRKNKHWPDLILTEAASLNSQ
GFWAYQFKKHGTCSSDLFNQEKYFDLALILKDKFDLLTFRNKGIIIPKSTCTINKIQK
TIRVTGVVPLNSCTPTMELLEVIGICFNDRDASKLIDCDQPKTCDTSGNTEIFFP"

BASE COUNT 683 a 390 c 444 g 724 t
ORIGIN

Alignment Scores:
Pred. No.: 51.7 Length: 2241
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 8 Gaps: 0

US-09-854-133-587 (1-16) x LPU28796 (1-2241)

QY 6 GlyTleAspPheIleIlePhe 12
|||||
Db 2203 GGAATTGATTTTATATATTT 2183

RESULT 23
AF043091/c
LOCUS
DEFINITION Hordeum vulgare dehydrin 6 (dhn6) gene, complete cds.
ACCESSION AF043091
VERSION AF043091.1 GI:4105110
KEYWORDS
SOURCE
ORGANISM
Hordeum vulgare subsp. vulgare.
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Hordeum.
1 (bases 1 to 3086)
Choi,D.-W., Zhu,B. and Close,T.J.
The barley (Hordeum vulgare L.) dehydrin multigene family:
sequences, chromosome assignments, and expression characteristics
of 11 dhn genes of cv. Dicktoo
Unpublished
REFERENCE
2 (bases 1 to 3086)
Choi,D.-W., Zhu,B. and Close,T.J.
Direct Submission
Submitted (15-JAN-1998) Botany and Plant Science, University of
California, Riverside, CA 92521, USA
Location/Qualifiers
1. .3086
/organism="Hordeum vulgare subsp. vulgare"
/cultivar="Dicktoo"
/db_xref="taxon:112509"
/chromosome="4H"
670. .>2395
/gene="dhn6"
join(670. .1892,1985. .>2395)
/gene="dhn6"
/product="dehydrin 6"
670. .674
/gene="dhn6"
TATA_signal
join(795. .1892,1985. .2395)
/gene="dhn6"
CDS
```

```
/note="Lea D-11 protein"
/codon_start=1
/product="dehydrin 6"
/protein_id="AAD02257.1"
/db_xref="GI:4105111"
/translation="MAHFQGGQHHGHPATRVDEYGNPVTAGGGGGVGTGDTGLGHPQGG
QHGRTRRLDEYGNPVTAGHVGAGSTGEVHGPCHAGYGSTGNTDGGHGRQVGYGA
TGTHDAGGSGGIAPRHGGAGTGVHDAGGLGHTTGHGATSTHGTGHTAGYGTGT
GMTGHTGHTAGYDATGTGTHGTGHTTGYGGATGTGTHGTGHTAGLGTGTGTG
GTGRTAGYGTGTGTHGTGHTTGYGGATGTGTHGTGHTAGLGTGTGTG
HGTGHTTGYGGVGTGTGTHGTGHTTGLGTTGTTGTTGTTGTTGTTGTTGTTG
THDVVGTGPHGGGLGEHKTGGILHRSGSSSSSSSSSSSSSSSSSSSSSSSSSS
GNKEQTATGGYGPYGTGTTGGAGHAGTGTGTTGTTGTTGTTGTTGTTGTTGTT
ATTGGYAGTTGTGGYAGTTGTGGTYGATEGTHEKKGMMEKIKEKLPGGH"

BASE COUNT 714 a 924 c 939 g 509 t
ORIGIN

Alignment Scores:
Pred. No.: 68.7 Length: 3086
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 8 Gaps: 0

US-09-854-133-587 (1-16) x AF043091 (1-3086)

QY 9 PheIleIlePheTrpIlePhe 15
|||||
Db 2801 TTTATTATATTTTGGATATTT 2781

RESULT 24
AX347114
LOCUS
DEFINITION Sequence 2185 from Patent WO0200928.
ACCESSION AX347114
VERSION AX347114.1 GI:18495002
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1
AUTHORS
Olek,A., Piepenbrock,C. and Berlin,K.
TITLE
Diagnosis of diseases associated with the immune system
JOURNAL
Patent: WO 0200928-A 2185 03-JAN-2002;
Epigenomics AG (DE)
FEATURES
source
1. .5886
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 1482 a 47 c 1292 g 3065 t
ORIGIN

Alignment Scores:
Pred. No.: 122 Length: 5886
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX347114 (1-5886)

QY 6 GlyIleAspPheIleIlePhe 12
|||||
Db 4386 GGGATTGATTTTATATATTTT 4406

RESULT 25
MITPCOCB
LOCUS
DEFINITION Theileria parva mitochondrion DNA.
ACCESSION 223263
```



VERSION 223263.1 GI:437862  
KEYWORDS cytochrome b; cytochrome oxidase; cytochrome oxidase subunit I;  
cytochrome oxidase subunit III; large subunit ribosomal RNA;  
ribosomal RNA.  
SOURCE Theileria parva.  
ORGANISM Mitochondrion Theileria parva  
Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileridae;  
Theileria.  
REFERENCE 1 (bases 1 to 5895)  
AUTHORS Kairo,A., Fairlamb,A.H., Gобрight,E. and Nene,V.  
TITLE A 7.1 kb linear DNA molecule of Theileria parva has scrambled rDNA  
sequences and open reading frames for mitochondrially encoded  
proteins  
JOURNAL EMBO J. 13 (4), 898-905 (1994)  
MEDLINE 94155854  
PUBMED 8112303  
REFERENCE 2 (bases 1 to 5895)  
AUTHORS Nene,M.  
TITLE Direct Submission  
JOURNAL Submitted (06-JUL-1993) Nene V. M., International Laboratory for  
Research on Animal Diseases Nairobi Kenya  
FEATURES  
source Location/Qualifiers  
1..5895  
/organism="Theileria parva"  
/organelle="mitochondrion"  
/isolate="Muguga"  
/db\_xref="taxon:5875"  
/germline  
1..60  
repeat\_region  
/partial  
/standard\_name="inverted terminal repeat"  
/citation=[1]  
/evidence=experimental  
/rpt\_type=TERMINAL  
38..1489  
/gene="coI"  
38..1489  
/partial  
/gene="coI"  
/function="electron transport"  
/note="The coordinates given is where an open reading  
frame starts. This ORF starts at the codon AGT."  
/citation=[1]  
/codon\_start=1  
/evidence=experimental  
/transl\_table=4  
/product="cytochrome oxidase subunit I"  
/protein\_id="CAA80798.1"  
/db\_xref="GI:437863"  
/db\_xref="SPTREMBL:Q36097"  
/translation="SFIKFGLFVFNVSNGNHKIIIGISYLLAWFWMGFYMSVLI  
RTELSMSGLKIITMDTLEIYNLLFTLHGLIMVEFNMTGLFGGIGNVLYPVLLIGYCDV  
VYPRVNLVSLFQPIGFVLVSSYLEIGSGTGWTLYPPLSTLSNVGIDFIIFGLLA  
AGIATLSSVNFITFTSVKTI GFVIDRISPAWSIVLTSFLLLLSLPVVTAFLMVF  
LDRHYNTMFFSSNSGDPVLYQHLFFGFHPEVYIMILPGFGIISLLSTYTTKEMFG  
NOTMILAMGNSFVRLFGLGTSYVHIRGSRYSRYFTTIVILIALPGNKIFENWVITL  
QCVESIKSLGLVLTFLFIVNFVIGTGTGVILGNAGIDLALHDTVYVVGHEHFFVLSIG  
AIIISMICFIYIQRMLLFGIILSNRLSLIAPIFMISVLLTFLPMHFTGFSPLPRRIP  
DYPDEMGWNFTCTLGATMMLVLKLAIFLISL"  
complement(2301..3068)  
/gene="coIII"  
complement(2301..3068)  
/gene="coIII"  
/function="electron transport"  
/citation=[1]  
/codon\_start=1  
/evidence=experimental  
/transl\_table=4  
/product="cytochrome oxidase subunit III"  
/protein\_id="CAA80799.1"  
/db\_xref="GI:437864"  
/db\_xref="SPTREMBL:Q36098"  
/translation="MRNSAQSYLKYINIINIFETLYLFYSTGLDTLEYIDSTYKNFII

MYVNOYLLYGTTLKYLVSGEFNMNSLTIFINSIREIMTSTTMVMYAIFGMFIFSEILV  
FSTFIWGYFHLRLSNPILAEINVEAYLQISDLVNTGSILVSIILHRVQESANFETDF  
FMEQLLLIGFIFLSLQNDXSLISLVVNNYWMYLYFFILTLGLHSLHVCAGGIFVLQIS  
YFYEGLDGSQRDEEFNAGVYWHFVEMIIWIALTMLLFLA"  
3093..3393  
/partial  
/product="large subunit ribosomal RNA (LSU1)"  
/note="Fragmented rRNA sequence; codes for part of domain  
IV and V of LSU rRNA."  
/citation=[1]  
complement(3404..3514)  
/partial  
/product="large subunit ribosomal RNA (LSU3)"  
/note="Fragmented rRNA gene; codes for part of domain V of  
LSU rRNA."  
/citation=[1]  
complement(4365..4403)  
/partial  
/product="large subunit ribosomal RNA (LSU2)"  
/note="Fragmented rRNA gene; codes for part of domain V of  
LSU rRNA"  
/citation=[1]  
complement(4423..5583)  
/gene="cob"  
complement(4423..5583)  
/gene="cob"  
/function="electron transport"  
/standard\_name="cob"  
/note="Partial cDNA sequenced; extends from nucleotide  
5506 to the end of the open reading frame - poly A  
addition on cDNA starts at nucleotide 4368."  
/citation=[1]  
/codon\_start=1  
/evidence=experimental  
/transl\_table=4  
/product="cytochrome b"  
/protein\_id="CAA80800.1"  
/db\_xref="GI:437865"  
/db\_xref="SPTREMBL:Q36099"  
/translation="MFYKKVIKIKLGNRALTYKLTIFIMNMENAHIPSYLPKNLNSNW  
NVGFIILGILLIQLISGLLLTFFYVPCKEGAFESLSRLVTETQGFVRLYHVSQVSE  
YFFFMFIHIKGMWYSSKYPWWSYGVILLSIVIAFTGYVLPGQMSFWGATVIS  
NLLEWFGKAKVITFGGFTVGPETLKRFFILHFVLPVAVLVIVLLHLFLHREGSSNPL  
TLAEAVALLKFYQLILFSDVKFLVIISMFIGVQYGYGWTFLQADNDNSILSSSENTP  
AHIPEWYLLLFATLKVPFTKVSGLVAMVVVKLLIILVESRSKQAVSTAHHRVW  
TTTSVPLVPALFLLCIGRMVINLDLIIIGIYGVLLSTTFVQKLLDSSRVRA"  
complement(5552..5619)  
/partial  
/product="large subunit ribosomal RNA (LSU5)"  
/note="Fragmented rRNA gene; codes for part of domain IV  
of LSU rRNA."  
/citation=[1]  
5691..5770  
/partial  
/product="large subunit ribosomal RNA (LSU4)"  
/note="Fragmented rRNA gene; codes for part of domain V of  
LSU rRNA."  
/citation=[1]  
complement(5783..5895)  
/partial  
/standard\_name="inverted terminal repeat"  
/note="Contains 12 bp direct and inverted sub-repeats."  
/citation=[1]  
/evidence=experimental  
/rpt\_type=TERMINAL  
BASE COUNT 2018 a 889 c 879 g 2109 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 122 Length: 5895  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0

```
Query Match: 43.75% Indels: 0
DB: 3 Gaps: 0
US-09-854-133-587 (1-16) x MITPCOCB (1-5895)
QY 6 GlyileAspPheIleIlePhe 12
Db 485 GGTATAGATTTTATAATATT 505

RESULT 26
AWU10505
LOCUS Acetobacterium woodii DSM 1030 Na+ translocating FlFO ATP synthase
DEFINITION linear BCT 22-NOV-1999
ACCESSION U10505 U63995
VERSION U10505.2 GI:6014710
KEYWORDS
SOURCE Acetobacterium woodii.
ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Eubacteriaceae;
Acetobacterium.
REFERENCE 1 (bases 5035 to 8237)
AUTHORS Forster,A., Daniel,R. and Muller,V.
TITLE The Na(+)-translocating ATPase of Acetobacterium woodii is a
FlFO-type enzyme as deduced from the primary structure of its beta,
gamma and epsilon subunits
JOURNAL Biochim. Biophys. Acta 1229 (3), 393-397 (1995)
MEDLINE 95267800
PUBMED 7748890
REFERENCE 2 (bases 1 to 8237)
AUTHORS Rahlfs,S. and Muller,V.
TITLE Sequence of subunit c of the Na(+)-translocating FlFO ATPase of
Acetobacterium woodii: proposal for determinants of Na+ specificity
as revealed by sequence comparisons
JOURNAL FEBS Lett. 404 (2-3), 269-271 (1997)
MEDLINE 97227959
PUBMED 9119076
REFERENCE 3 (bases 1 to 8237)
AUTHORS Rahlfs,S. and Muller,V.
TITLE Sequence of subunit a of the Na(+)-translocating FlFO-ATPase of
Acetobacterium woodii: proposal for residues involved in Na+
binding
JOURNAL FEBS Lett. 453 (1-2), 35-40 (1999)
MEDLINE 99330164
PUBMED 10403370
REFERENCE 4 (bases 1 to 8237)
AUTHORS Rahlfs,S., Aufurth,S. and Muller,V.
TITLE The Na(+)-F(1)F(0)-ATPase operon from Acetobacterium woodii. Operon
structure and presence of multiple copies of atpE which encode
proteolipids of 8- and 18-kDa
J. Biol. Chem. 274 (48), 33999-34004 (1999)
MEDLINE 20036535
PUBMED 10567365
REFERENCE 5 (bases 5035 to 8237)
AUTHORS Muller,V.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1994) Institut fuer Mikrobiologie, Goettingen,
37077, Germany
REFERENCE 6 (bases 1 to 8237)
AUTHORS Muller,V. and Rahlfs,S.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1999) Lehrstuhl fuer Mikrobiologie, University of
Munich, Maria-Ward-Str. 1a, Munich 80638, Germany
REMARK Sequence update by submitter
COMMENT On or before Oct 7, 1999 this sequence version replaced gi:4713917,
gi:501146.
FEATURES
source Location/Qualifiers
1..8237
/organism="Acetobacterium woodii"
/strain="DSM 1030"
/db_xref="taxon:33952"
1..8237
/gene="Na+ translocating FlFO ATP synthase operon"

gene
CDS
```

```
/note="The ATPase of Acetobacterium woodii is of special
interest because it uses
sodium ions instead of protons as physiological coupling
ion and it
contains multiple genes encoding 8- and 16-kDa
proteolipids."
257..658
/gene="atpI"
257..658
/gene="atpI"
/codon_start=1
/transl_table=11
/product="Unci homolog"
/protein_id="AAF01472.1"
/db_xref="GI:6014712"
/translation="MKTMKIFNKLSLETKIMIDSGLISVGLAFSLTQNPLTFALGV
LFGGLYSILNFKLMQLTFDRKMKPSARAQKYVQTRYFLRYLITGVVIYVAIMNPWN
IIGVLLGLVAVKISVLENTTLPKKNASINEA"
724..1386
/gene="atpB"
724..1386
/gene="atpB"
/codon_start=1
/transl_table=11
/product="FlFO ATPase a subunit"
/protein_id="AAF01473.1"
/db_xref="GI:6014713"
/translation="MEGPKIYGELFGLPITQTLVNTWIIMALILVVCYILTKNMQIKR
PGKAQVIAETIVTALDLDLVGQTMGDKMKFAPYMLALIMELAFSNTAGIYGLRSPTAD
LNCTVACFDDVFEDAVLWFKIKRQLSQGFPEPMPFLPINVIGEFANPVSLSFRLF
GNMLGGGIIMALLYALAGLIFFPVLIPVPFHFFYDFISGLLSQSFIFMTMLTMVFISGA
MD"
1410..1958
/gene="atpE1"
1410..1958
/gene="atpE1"
/note="16-kDa proteolipid homolog"
/codon_start=1
/transl_table=11
/product="FlFO ATPase c1 subunit"
/protein_id="AAF01474.1"
/db_xref="GI:6014714"
/translation="MIGDMNIVDFVIQFLSQFDPVDVIKGFSAIGLAMVAGVGPGI
GQGFAGKGAEEAVGKNPTKSNIDIVMIMLGAAVAETSGIFSLVIALILLFANPFISST
ASVWILSASAMASGAMIAGIGPGTGQGYAAGKGAEEAVGIRPEMKSAILLRVMLLGQAV
AQTGTIYALIVAILMYANPFL"
2001..2249
/gene="atpE2"
2001..2249
/gene="atpE2"
/note="8-kDa proteolipid homolog"
/codon_start=1
/transl_table=11
/product="FlFO ATPase c2 subunit"
/protein_id="AAC45088.2"
/db_xref="GI:4713918"
/translation="MEGLDFIKACSAIGAGIAMIAGVGPVGIGQGFAGKGAEEAVGRQP
EAQSDIIRTMLLGAAVAETTGIIYGLIVALLFANPFF"
2306..2554
/gene="atpE3"
2306..2554
/gene="atpE3"
/note="8-kDa proteolipid homolog"
/codon_start=1
/transl_table=11
/product="FlFO ATPase c3 subunit"
/protein_id="AAF01475.1"
/db_xref="GI:6014715"
/translation="MEGLDFIKACSAIGAGIAMIAGVGPVGIGQGFAGKGAEEAVGRQP
EAQSDIIRTMLLGAAVAETTGIIYGLIVALLFANPFF"
2572..3126
/gene="atpF"
2572..3126
```

```
/gene="atpF"
/codon_start=1
/transl_table=11
/product="F1FO ATPase b subunit"
/protein_id="AAF01476.1"
/db_xref="GI:6014716"
/transl_table=11
/translation="MKNLNAYKLERGIYPLIFEYAGLVGIDLKLLLATIVNFIIFELL
LEHFFYGVKVDILAKRQDDVTAEIVGATEKNTAAAKLQOEYGLSDIRAKEREIIRN
ATVEGQSERQEIIDKAHSDAKLIIIEKAMAEIEVEKRAMNEVKSNIIVDLSLYAAEKII
VETLAQKKHEAMVLDIFDKGDAQ"
3123.3665
/gene="atpH"
3123.3665
/gene="atpH"
/codon_start=1
/transl_table=11
/product="F1FO ATPase delta subunit"
/protein_id="AAF01477.1"
/db_xref="GI:6014717"
/transl_table=11
/translation="MSLVASKYARALFDVAVDKDQDLEIFSDFKTATDLFSSEKFKMD
IMLTPLSLNTGKGGKILMRSLSLSNQYVKNYLMILMDKNRFEDIFDIYEAFRKLCNEH
KNLVEARVLTIVPLDETLLRIALEENLAKRFNKKVILENEIDKLSILGAVVYVGDQIID
GSIKQLSQMKKQVNNLRH"
3682.5187
/gene="atpA"
3682.5187
/gene="atpA"
/EC_number="3.6.1.34"
/codon_start=1
/transl_table=11
/product="F1FO ATPase alpha subunit"
/protein_id="AAA79906.2"
/db_xref="GI:6014711"
/transl_table=11
/translation="MNLREEISQIIKNEIDRYEDKLEVDVGTVIQVGDGVARVHGL
ENAMAGELLAFNEVYGMVLNLEEDNVCVLLGYDDDIVGDIIVRCTGRILKFLQLANH
VFGRVNMGALGFPVQDGPVIVDHRPVEVKAAGVIERESVNQPIQTGYKALDSMIPIG
RGQRELIIGDRQTKTALGFDTIINQKGEDVICIYVAIGOKDSTVAQIVGQLEENNAM
DYTIIVSAGAAQLAPLHNIAPYSGVTMAEYFMKGPKRTIVHQGRPSKHGCLPSLVFD
PRPPGPEPYPGDVFLHSLRLLERAAKLKAGSITALPIIETOAGDVSAIYPTNVISI
TGQIFLEAEFRSGIRPAVNPGISVRSVGSQAQIKSMKKVAGPLRIEYQVRELASF
AQFGSLDDETKAQLAKGERIVEILKQDQYDPMNVEDQVLILYAATNGFLLDIEVKDI
REFEGLIKFAQKKYPEIMTKVKKGDLSDLEVVAFAECIEAYKKVFSKV"
5212.6114
/gene="atpG"
5212.6114
/gene="atpG"
/EC_number="3.6.1.34"
/codon_start=1
/transl_table=11
/product="F1FO ATPase gamma subunit"
/protein_id="AAA79907.1"
/db_xref="GI:501148"
/transl_table=11
/translation="MAENVQDIKPRIKSVNSTMQITHAMELVASAKLRKSRELAEGRR
PYFEAMIESIGRIVEKSGNARNIFMDQREVKKTYIITGDKGLAGGYNVNVAKLVEE
HITDKENAVLFTVSGSRGRDHFNRREYHIQGEYLISERPNNFFNAKEVTAIVMEGKNG
EYDEVYIATKRVSTITQHAQMKLLPLSREELITSGKVKTTEETKEEKSMSDRELT
IMTYEPEPEELKYLIPNLVSVSTVYGSIMESAAEQGARRTAMESATTNANEMIDGLT
LQYNRVRAPIITQBISEIVGGAED"
6218.7618
/gene="atpD"
6218.7618
/gene="atpD"
/EC_number="3.6.1.34"
/codon_start=1
/transl_table=11
```

Alignment Scores:  
Pred. No.: 165  
Score: 7.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 43.75%  
DB: 1

Length: 8237  
Matches: 7  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0

```
US-09-854-133-587 (1-16) x AWU10505 (1-8237)
QY 5 CysGlyIleAspPheIleIle 11
Db 2210 TGTGSCATGATTATTATT 2230
RESULT 27
AX347131 LOCUS 8576 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 2202 from Patent WO0200928.
ACCESSION AX347131
VERSION AX347131.1 GI:18495019
KEYWORDS synthetic construct.
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 2202 03-JAN-2002;
Epigenomics AG (DE)
FEATURES
source
1.8576
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2122 a 64 c 1723 g 4667 t
ORIGIN
Alignment Scores:
Pred. No.: 171 Length: 8576
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
Gaps: 0
DB: 6
US-09-854-133-587 (1-16) x AX347131 (1-8576)
QY 6 GlyIleAspPheIleIlePhe 12
Db 5203 GGAATAGATTTTATAATATT 5223
RESULT 28
AX339183 LOCUS 9265 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 50 from Patent WO0176451.
ACCESSION AX339183
VERSION AX339183.1 GI:18129260
KEYWORDS synthetic construct.
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with metabolism
JOURNAL Patent: WO 0176451-A 50 18-OCT-2001;
Epigenomics AG (DE)
FEATURES
source
1.9265
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2686 a 133 c 1801 g 4645 t
ORIGIN
Alignment Scores:
Pred. No.: 183 Length: 9265
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
```

```
DB:
US-09-854-133-587 (1-16) x AX339183 (1-9265)
QY 9 PheillePheTrpIlePhe 15
Db 525 TTTATATTTTGGATATTT 545

RESULT 29
AX347055
LOCUS AX347055
DEFINITION Sequence 2126 from Patent WO0200928.
ACCESSION AX347055
VERSION AX347055.1 GI:18494943
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
    source
        Location/Qualifiers
            1..9265
                /organism="synthetic construct"
                /db_xref="taxon:32630"
                /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2686 a 133 c 1801 g 4645 t
ORIGIN

Alignment Scores:
Pred. No.: 183 Length: 9265
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX347055 (1-9265)
QY 9 PheillePheTrpIlePhe 15
Db 525 TTTATATTTTGGATATTT 545

RESULT 30
AE001096
LOCUS AE001096
DEFINITION Archaeoglobus fulgidus section 11 of 172 of the complete genome.
ACCESSION AE001096 AE000782
VERSION AE001096.1 GI:2689419
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
```

REFERENCE  
AUTHORS

2 (bases 1 to 11046)

Klenk,H.P., Clayton,R.A., Tomb,J.-F., White,O., Nelson,K.E., Ketchum,K.A., Dodson,R.J., Winn,M., Hickey,E.K., Peterson,J.D., Richardson,D.L., Kerlavage,A.R., Graham,D.E., Kyrpides,N.C., Fleischmann,R.D., Quackenbush,J., Lee,N.H., Sutton,G.G., Gill,S., Kirkness,E.F., Dougherty,B.A., McKenney,K., Adams,M.D., Loftus,B., Peterson,S., Reich,C.I., McNeil,L.K., Badger,J.H., Glodek,A., Zhou,L., Overbeek,R., Gocayne,J.D., Weidman,J.F., McDonald,L., Utterback,T., Cotton,M.D., Spriggs,T., Artiach,P., Kaine,B.P., Sykes,S.M., Sadow,P.W., D'Andrea,K.P., Bowman,C., Fujii,C., Garland,S.A., Mason,T.M., Olsen,G.J., Fraser,C.M., Smith,H.O., Woese,C.R. and Venter,J.C.

## TITLE

## JOURNAL

## REMARK

## COMMENT

## FEATURES

## source

Location/Qualifiers

1..11046

/organism="Archaeoglobus fulgidus"

/db\_xref="taxon:2234"

91..876

/gene="AF0140"

91..876

/gene="AF0140"

/note="similar to GB:M87049 SP:P27851 PID:148231 GB:U00096  
PID:2367307 percent identity: 31.00; identified by  
sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="ubiquinone/menaquinone biosynthesis  
methyltransferase (ubiE)"

/protein\_id="AAB91083.1"

/db\_xref="GI:2650497"

/translation="MITVLILRCIMVKSTTRKSAKSDRRRYNRIARFYDLFETPMELM  
NFKWRKMLFDLIIKAEGVLVEICAGTKNIPHSNVVALDISEKMLERAVRRAKES  
GKKVDFLLADAENLPFRNSFDVFTFVFCSDVDDPVRGLREAFVLKGGRAYFLEH  
MLPKSRFLHPLNLLNPIFAFGPEINRRTDENIQAGFIQVKEYMLFGSVFRLIVAE  
KPLDSVDEASGYSLSKRSSESSDNFRFFIGQSLPLRLTQLLR"

## gene

/gene="AF0141"

883..1182

/gene="AF0141"

/note="hypothetical protein; identified by GeneMark;  
putative"

/codon\_start=1

/transl\_table=11

/product="A. fulgidus predicted coding region AF0141"  
/protein\_id="AAB91091.1"

/db\_xref="GI:2650507"

/translation="MNWKKVGIIVAIATAAGLYVAAAMPWGMGYGPMGYTPMMGYGG  
GMGMGMGYAPQYGPQGYGPAYNYTDTTPQPPQGYVYPYGGHGCMPMGYW"

## gene

/gene="AF0142"

1200..1661

/gene="AF0142"

/note="similar to GB:L09121 PID:155085 SP:P98052 percent  
identity: 38.04; identified by sequence similarity;  
putative"

/codon\_start=1

/transl\_table=11

/product="cytochrome C oxidase, subunit II, putative"  
/protein\_id="AAB91082.1"

/db\_xref="GI:2650496"

/translation="MRKVLTPIFLILIPAIISQGNLLSPQAGVKNKTVYIEAFQFG  
YRVVAIQEGDKLYPSTPSLKLNAGDVTTFHVTVKDVTHGFTIDGITDGAIELLVIP  
GRTLTVGPVTFDVGKRVRCNSANGSLHFPMSIDIIVKIPYALSALL"

## gene

/gene="AF0143"

1661..3070

/gene="AF0143"

1661..3070



/gene="AF0143"  
/note="similar to GB:L77117 PID:1592228 percent identity:  
25.53; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAB91081.1"  
/db\_xref="GI:2650499"  
/translation="MVGFAALLYDGGGEIRFRARLLKPRWVIFLLTAINLSVFLLI  
IFTGLFGSAVGSSENFSAIVTILWFALVEFMILFAARVWCACVPIPIVGEWLKRRRIL  
GVREFESFTFPARYRNWIPAIAGLLLSLEIPWIVTKPLVTALLFISLAVATAMYL  
IYPPRYFCRHVCASGYIGHSNASLLALRSEDEKCRCEKFKLGSEGYACPWKL  
YPGNEENTYCGLECLRTCRDNLNLSRLRPFEELSKKA VRAGKALKLDEAAMGFI  
RFLAIFYELFFGSLYWLKDFGNMGNPNYGVLENAYLLLPSTFGIVKWLWAGIIVVG  
TIAAFAGLYYIFSWIGARIAKVDTKQLFVALSYSLAPYGLLVWAGFAVGLLAYFHIYP  
FSAFLDPFGVEKNAPVANVKLLIIQIALFIAAGLASAIVATKRVAKRLTDKFEALTA  
VYVPHLIAALTILIKIAG"  
3075. .3455  
/gene="AF0144"  
3075. .3455  
/gene="AF0144"  
/note="similar to GB:L09121 PID:155085 SP:P98052 percent  
identity: 34.18; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome C oxidase, subunit II (cbaB)"  
/protein\_id="AAB91084.1"  
/db\_xref="GI:2650498"  
/translation="MKKEIAALLFVLISAI PAYLTYDYIESRKEKNPEIVIEAHIT  
ESGGFKPNYIVVPKSKVKLVFKAADVTHGIFIPSLGINELIPAGKEVVVEFYAEKEG  
VYPFMCVYCSPIYHFRMGIVV"  
complement(3481. .3819)  
/gene="AF0145"  
complement(3481. .3819)  
/gene="AF0145"  
/note="hypothetical protein; identified by GeneMark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="A. fulgidus predicted coding region AF0145"  
/protein\_id="AAB91090.1"  
/db\_xref="GI:2650505"  
/translation="MCRCRCMAMMSFVAIVTPLTMTGLIDWKYRYDMRKVPPIIQKV  
ITGIVGYVFWVYVVLHSLTDYSLAALAMALVFFAITGEYGGKLVHGARTLLCLKNLR  
KGSSQKSNPN"  
complement(4030. .4245)  
/gene="AF0146"  
complement(4030. .4245)  
/gene="AF0146"  
/note="hypothetical protein; identified by GeneMark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="A. fulgidus predicted coding region AF0146"  
/protein\_id="AAB91092.1"  
/db\_xref="GI:2650506"  
/translation="MIIAIVAVIFLLNFLTPTGYMPMGYGGMMGYSDERYPPGC  
CQGAESSSSNGSMITKPTGTACQGR"  
complement(4308. .4832)  
/gene="AF0147"  
complement(4308. .4832)  
/gene="AF0147"  
/note="similar to GB:L77117 SP:Q58204 PID:1592308 percent  
identity: 38.55; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical transmembrane protein"  
/protein\_id="AAB91080.1"  
/db\_xref="GI:2650495"  
/translation="MLKLSQALPESLKSFFIAGAYIQLVSTFLGFFAAWLIITAAVMHG  
LSAFFDGRGELRRTFEFAGYGFMPSLIGSAVTPVSLKYVEATIPTIDLQELSANPE

ILVKSLVSHFPDSYVYSAIFLNLA VTAWSFLIWTFA LKNARNVELKQAAVCAAIPTAI  
FGLHQVMALVRLIQ"  
5068. .5325  
/gene="AF0148"  
5068. .5325  
/gene="AF0148"  
/note="hypothetical protein; identified by GeneMark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="A. fulgidus predicted coding region AF0148"  
/protein\_id="AAB91089.1"  
/db\_xref="GI:2650504"  
/translation="MEARGWDEEVMKWLAVAICLAMVGMVMPAFQPLNLA FELYYG  
HHESLPITAASAAAYEGIVITATLAAAAATAELVHLLLQQFL"  
5470. .5724  
/gene="AF0149"  
5470. .5724  
/gene="AF0149"  
/note="hypothetical protein; identified by GeneMark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="A. fulgidus predicted coding region AF0149"  
/protein\_id="AAB91088.1"  
/db\_xref="GI:2650503"  
/translation="MVFAVYLVSGFMLVLGIKYGRTKIYSVKAGLILAGFAVIFAA  
AWIAFTGTSASQERIGLAKSLSVVMGLIAGVLSVYVLSKS"

Alignment Scores:  
Pred. No.: 214 Length: 11046  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 1 Gaps: 0

US-09-854-133-587 (1-16) x AE001096 (1-11046)

QY 6 GlyIleAspPheIleIlePhe 12  
|||||  
Db 4971 GGCATCGATTATTAATATTT 4991

RESULT 31  
AE007643/c  
LOCUS AE007643 11995 bp DNA linear BCT 27-JUL-2001  
DEFINITION Clostridium acetobutylicum ATCC824 section 131 of 356 of the  
complete genome.  
ACCESSION AE007643 AE001437  
VERSION AE007643.1 GI:15024237  
KEYWORDS  
SOURCE  
ORGANISM  
Clostridium acetobutylicum.  
Clostridium acetobutylicum  
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
Clostridium.

REFERENCE  
AUTHORS  
1 (bases 1 to 11995)  
Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,  
Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,  
Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,  
Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.  
Genome sequence and comparative analysis of the solvent-producing  
bacterium Clostridium acetobutylicum  
J. Bacteriol. 183 (16), 4823-4838 (2001)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
2 (bases 1 to 11995)  
Childress,D., Zeng,Q. and Smith,D.R.  
Direct Submission  
Submitted (24-JUL-2001) GTC Sequencing Center Production, 100  
Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100  
Beaver Street, Waltham, MA 02453-8443, USA  
FEATURES  
Source  
1. .11995  
Location/Qualifiers

gene /organism="Clostridium acetobutylicum"  
CDS /strain="ATCC 824"  
/db\_xref="ATCC:824"  
/db\_xref="taxon:1488"  
90..239  
/gene="CAC1307"  
90..239  
/gene="CAC1307"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="AAK79278.1"  
/db\_xref="GI:15024238"  
/translation="MDYVFEDELKILAKDPNVKTVEDVQNILKSCLEKICFSRCSKE  
NWTNI"  
698..1291  
/gene="CAC1308"  
698..1291  
/gene="CAC1308"  
/codon\_start=1  
/transl\_table=11  
/product="Uncharacterized low-complexity protein, related  
to YISX/YYBG B.subtilis"  
/protein\_id="AAK79279.1"  
/db\_xref="GI:15024239"  
/translation="MSYNIDNKFELHHRWVETSGHEGEKLCLEEDVDFRKFDSLDKLFE  
QANLAECVFDNLNLENIDHTSLCSSTYKTNLRGCDYKTDLSYDFSDAVIKNR  
FSKSDCWESIFRNVNLINLNDYFYLTDFSNATLDNVDISLSTFEETLLNGVKLRN  
IKGIEEAHFESINIGTFEKPILLKSQEAKKVLENCI"  
complement(2210..2704)  
/gene="CAC1310"  
complement(2210..2704)  
/gene="CAC1310"  
/codon\_start=1  
/transl\_table=11  
/product="Integrase XerD family protein (similarity only  
with C-term. part"  
/protein\_id="AAK79280.1"  
/db\_xref="GI:15024240"  
/translation="MINTLTIKNNLIYTVGVKMSPPKTIESQRSVYISNFILSLKT  
LHKQLENKSNKAEFYKDNMFDGKRYDLMTWCSGKIYHPMLYTNKIKKVLKAAKINK  
NVRFHDLRHTNATLLQQGTDLKVIQERLGHKDITATANIYSHVNKSMQKAATEKLLN  
ILKD"  
3281..>3751  
/gene="CAC1311"  
/note="possible gene fragment 'Hypothetical protein, CF-36  
family"  
4007..4267  
/gene="CAC1312"  
4007..4267  
/gene="CAC1312"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="AAK79281.1"  
/db\_xref="GI:15024241"  
/translation="MAEISIPEDIEYDDSSMVSLYSNSVRMETSIIYDLQFIFESQYS  
KNGKPTIKQLARVIMSPLANMLNRNIFEYKKFGEIKLPNEL"  
4285..4905  
/gene="CAC1313"  
4285..4905  
/gene="CAC1313"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="AAK79282.1"  
/db\_xref="GI:15024242"  
/translation="MNKKQKEINNSIYIEDSTHNSILYTQEFVEELFHEGYKLGFLQG  
TTYANKIFKSKSTVDLNMVEKLSVEIEGIANKLNLDIIINQNKLTDNLDNLSSR  
VNDSSKTLMNKIDLSHRVEGIDKRTIVLKKSLDYIKEDICDNKSSIKDIKEDTKDL  
SIIRNNIGNMDEKSKWKQFLAPIIISATLSGLVLAIMLCIKSIK"  
complement(5011..5484)  
/gene="CAC1314"

CDS complement(5011..5484)  
/gene="CAC1314"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="AAK79283.1"  
/db\_xref="GI:15024243"  
/translation="MKFNKITSLLIAFSIIIGTSLSVSATTMAKTNKSNNTNIQSKLYD  
SETNLLNPTIDLNGTISWNPVNVVAYYFTTISKEKTFNNGTFATPVFNGLIHDAPTSI  
SFGADIYSNYGFGTYFAEVNAFDKNGSLQRLDISVYYDSDYGGPKLTKEFPNS"  
complement(5498..5944)  
/gene="CAC1315"  
complement(5498..5944)  
/gene="CAC1315"  
/codon\_start=1  
/transl\_table=11  
/product="Peptidoglycan-binding domain containing protein"  
/protein\_id="AAK79284.1"  
/db\_xref="GI:15024244"  
/translation="MKNLKLKALSLGIVSTVLISSSAFAAIPTVIQNHPIPKVAPVA  
TNLNATKSKMFASRTYGTSGRYMTAQQVIDNGGIIQEGDTGEPVADIQRRLGIYDD  
GIFGSATYSAVVDFQRRIDREFHRNGYSDPLALDGIQAQTWYLYK"  
6351..6704  
/gene="CAC1316"  
6351..6704  
/gene="CAC1316"  
/codon\_start=1  
/transl\_table=11  
/product="Predicted membrane protein"  
/protein\_id="AAK79285.1"  
/db\_xref="GI:15024245"  
/translation="MKKVGIIISIVFLVLSFQSCAAGVNALEKSSSSGSAGFILA  
VCMLIAGIIAIVSKYSKGMTITSIVFYAIGGIIGIANVGTYKDIQIWSILSIFAALI  
IFHIVKNKELYTGTK"  
complement(6878..7648)  
/gene="CAC1317"  
complement(6878..7648)  
/gene="CAC1317"  
/codon\_start=1  
/transl\_table=11  
/product="Potassium channel subunit"  
/protein\_id="AAK79286.1"  
/db\_xref="GI:15024246"  
/translation="MNITTHNKILYELTSSLALIASIMLIIELSVKLSNETLYIFDT  
IDNIIIFIITIDYFLRLYISKDKKFFKENIIDLSIIPFNSIFQGRLLKISKLLKF  
TKLIKFLKLFMAFSLRLRFKSHFTKFTKNPHYVYITTFVLITGTIGMHLTEGLSL  
GNALWSPVITITVGYGDISPSTPFGRVIASILMIGIGFLSMLTGTISTFFISKNG  
ISYRNEIENIKSKLDNDELSTEDINDICSILKTLKN"  
8086..8598  
/gene="CAC1318"  
8086..8598  
/gene="CAC1318"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="AAK79287.1"  
/db\_xref="GI:15024247"  
/translation="MKKKVYKNRLALFLSAIIILTSFVGCSESSSTSKSNSNTAKTIA  
KPKPKTAQSVLDELKKKESAYMTNITVTDNDSNKLGRPNQYTEKVTWKDNRSKDS  
QVDCSIELFKNKDDATARKDYLEIKIISMPMLTQYIEQKDNVLLRIDGVLTPEQSKEY  
INIFNSIFKN"  
8988..9689  
/gene="CAC1319"  
/note="glpF"  
8988..9689  
/gene="CAC1319"  
/codon\_start=1  
/transl\_table=11  
/product="Glycerol uptake facilitator protein, GLPF"  
/protein\_id="AAK79288.1"  
/db\_xref="GI:15024248"  
/translation="MTIFFAELVGTLLILLGDGVVANVVLKNSKGHASGWIVITMGW  
AFAVAIPALIFGSYNSQNFNPALTIALAVIGKVAWAKVPIYLAGQFIGAFLGAVLVFVV

YVDKFCSENKTKLGVCTPAVRNSLINFELCEVIGTEVLVFGILGVGAQNKNMG  
AVFVGLILVIGLSLGGPTGYAINPARDLAPRIAHLVLPNKGDSNWSYAWIPIIAP  
VVGIGIACVCMLLL"  
9766. .10323  
/gene="CAC1320"  
/note="glpp"  
9766. .10323  
/gene="CAC1320"  
/codon\_start=1  
/transl\_table=11  
/product="Glycerol-3-phosphate responsive antiterminator  
(mRNA-binding), GLPP"  
/protein\_id="AAK79289.1"  
/db\_xref="GI:15024249"  
/translation="MNIRELLEENPVIAVKNQEQQLAVNSEIQIVFLFGDIVSIK  
EISKLIKSKNIGIHLIDLEGLTNKEVIRYIKETEFDGIISTKSQIVKNAIKHNL  
IAVQRVFIFDTISLNNVKSHIIECDIAIEVLPGVIPKVLQIIAKHSNKPVVAGGLIET  
KEEVIEALKYGATCVSTTRKEIWM"  
10354. .11850  
/gene="CAC1321"  
/note="glpk"  
10354. .11850  
/gene="CAC1321"  
/codon\_start=1  
/transl\_table=11  
/product="Glycerol kinase, GLPK"

Alignment Scores:  
Pred. No.: 231 Length: 11995  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 1 Gaps: 0

US-09-854-133-587 (1-16) x AE007643 (1-11995)

QY 6 GlylleAspPheillePhe 12  
DB 8480 GGCATAGATTTTATTATTATT 8460

RESULT 32  
AE001070/c 12971 bp DNA linear BCT 15-DEC-1997  
LOCUS Archaeoglobus fulgidus section 37 of 172 of the complete genome.  
DEFINITION Archaeoglobus fulgidus section 37 of 172 of the complete genome.  
ACCESSION AE001070 AE000782  
VERSION AE001070.1 GI:2689393

KEYWORDS Archaeoglobus fulgidus.  
SOURCE Archaeoglobus fulgidus  
ORGANISM Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
Archaea; Euryarchaeota; Archaeoglobus.  
Archaeoglobaceae; Archaeoglobus.

REFERENCE 1 (bases 1 to 12971)  
AUTHORS Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E.,  
Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,  
Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C.,  
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,  
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,  
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,  
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,  
Utterback, T., Cotton, M.D., Spriggs, T., Artach, P., Kaine, B.P.,  
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,  
Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,  
Woese, C.R. and Venter, J.C.

TITLE The complete genome sequence of the hyperthermophilic,  
sulphate-reducing archaeon Archaeoglobus fulgidus  
JOURNAL Nature 390 (6658), 364-370 (1997)  
MEDLINE 98049343  
PUBMED 9389475

REFERENCE 2 (bases 1 to 12971)  
AUTHORS Klenk, H.P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E.,  
Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,  
Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C.,

Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,  
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,  
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,  
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,  
Utterback, T., Cotton, M.D., Spriggs, T., Artach, P., Kaine, B.P.,  
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,  
Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,  
Woese, C.R. and Venter, J.C.

TITLE Direct Submission  
JOURNAL Submitted (15-DEC-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
REMARK In order to show the genes in ascending order on the genome, the  
origin of this version has been moved by TIGR to position 2093570  
of the original version and the opposite strand is shown from the  
original version.

On Dec 16, 1997 this sequence version replaced gi:2650129.

COMMENT  
FEATURES  
source

1. 12971  
/organism="Archaeoglobus fulgidus"  
/db\_xref="taxon:2234"

gene

CDS

80. .295  
/gene="AF0483"

80. .295  
/gene="AF0483"

/note="similar to GB:L77117 PID:1500013 percent identity:  
32.26; identified by sequence similarity; putative"

/codon\_start=1  
/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAB90752.1"

/db\_xref="GI:2650141"

/translation="MKRILVCPVCKSEVELDAGYTGKYYCKNCGYVGSFILEMTEG  
EYREMEKEKFERKEDEKSKPKGVRED"

329. .1465  
/gene="AF0484"

329. .1465  
/gene="AF0484"

/note="similar to PID:1212775 SP:P51695 percent identity:  
44.53; identified by sequence similarity; putative"

/codon\_start=1  
/transl\_table=11

/product="GTP cyclohydrolase II (riba-1)"

/protein\_id="AAB90751.1"

/db\_xref="GI:2650140"

/translation="MEEVSSHVKSGKPMILDSKSAICIPAENVSGEILNFMKNCD  
EIRLALTWKQILNLGNLFRFRQNGNLIPIPNLEKISAEERAKFIRELVSGNAGDIKY  
PGRIFVEETKEMGLERPGIAEACVDLARMAGFAPSAYAPLMTAEGGVAGEDYALKF  
AKEHMPVFRIKDMIEFRIKSEKIVERVIEATLPTKFGTFRVAGYKTPGLGEIVALKV  
GRVDEGDLVRIHSECLTGDVDFHSLRCDGQDQLENALKMIDREGKGVAIYMRGHEGRG  
IGLNLKLMAYKLOEGKTDVDANIELGFPDPMRSYGAIAQILMDLKVKSIRLLTNPL  
KIEELKKYGFKIVREPIEVEPCEVNLPLYLKAKKDKMGLICFND"

1450. .2187  
/gene="AF0485"

1450. .2187  
/gene="AF0485"

/note="hypothetical protein; identified by GeneMark;  
putative"

/codon\_start=1  
/transl\_table=11

/product="A. fulgidus predicted coding region AF0485"

/protein\_id="AAB90754.1"

/db\_xref="GI:2650143"

/translation="MLQRLICEGAELSSCAAFCLNCGRRHVGCGVYVSESRRFVRI  
FGERDFEESIKRYDEDSIRNLYEILAERIYSYRVESIVVSGSEELIDEALEGLRS  
TLYPFDVSDAFDSQPSFFERLERVLRLEKELTKVIPPEEKIHGSHSTIIGREGY  
KLILSLAKSPYKKVVPGVIEGNATSIGGGVKLKLTRSDKGNVRALLIDGSSVQQVY  
VITASSREEGEVLLKLSGYVRDLQE"

2165. .2551  
/gene="AF0486"

2165. .2551  
/gene="AF0486"

/note="similar to GB:L77117 SP:Q58075 PID:1591374 percent  
identity: 28.92; identified by sequence similarity;

gene

CDS

gene

CDS

putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAB90750.1"  
/db\_xref="GI:2650139"  
/translation="MYAIYRSEGVIEVAIPFECRMCGKCEKLSQCVDPSPKGIIEA  
DLGLFEYVELNDVELKSPILIKPCPLKDNRCIAHEIRPKSCREFPLTERGDMGVGCE  
GYREIRAKAKAVEKALGKLEFAGRLD"  
2609. .4942  
/gene="AF0487"  
2609. .4942  
/gene="AF0487"  
/note="similar to SP:P31956 PID:142210 percent identity:  
31.71; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="penicillin G acylase"  
/protein\_id="AAB90749.1"  
/db\_xref="GI:2650138"  
/translation="MKKLYAAIPLAVLVLLVLAHLYLNLAPFSGSLWEIDISKEIP  
NPYGAEVYFDEYGVPHIVAENEKALAFAGVYHAKDRLFQMDLHRRMMKGQLSEVFG  
EDFDSDEFHVKMDFLGAAEASWSVLKDSIGDLIVAYSEGVMYIKNRPLPMEFKLA  
GYKPKETPVDLSLLGKEISWSLTGRFWDLSERFIVEKLGKALELYPDYLNHSYPII  
RTGMVNSLLEWLKPFKADGFGSNWVSGKLTANGKPLANDPHLLMLAPPVWYEM  
HLKVGNGENVGSLPGVPVILIGMNDYVANGFTNVGADVIDFYIYVWNGDKLYKGEM  
LEPEKSIKVKTKKKGIEERQVQVVEKTVHGVLEKYGRKIAVAWTGLTATTEALAVY  
KYNHAHNISEFMELGLKLFDPVQNVVADIYGNVMYYPAGKFPRLVNGTEVPGNVIF  
NGSAGEGWRGFKPYGISWDDGFIPFEEIPLINPDYVGTANQVVFQYHVLGDSMY  
FADPYRGMRIYEMLDEAANSGEKITPEYFMKMRDQVYSKPAEFFTPFIEAEKMSDK  
ARSAEKLNRNDFRMQKDSAAALFISWLDNFVNETFGDEFYSAGLESSYPHLYVLQ  
NLPSNSIYDDIRTEQREQRDDIAARAMEKAVEIIEETKGYRVYGDYNRLEIKHAFSSF  
VKFFDPSMPMNGSTYTVFNFRRAVDWGTGLSQAGSSWRMIVNFENYCVIPGGNSGN  
YFSPHYVDQLEMWANGQYKSFDFNVRGEKIVREVSS"  
4939. .5259  
/gene="AF0488"  
4939. .5259  
/gene="AF0488"  
/note="hypothetical protein; identified by GeneMark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="A. fulgidus predicted coding region AF0488"  
/protein\_id="AAB90755.1"  
/db\_xref="GI:2650144"  
/translation="MKRLIFVIGVFIVALALSFAHWVGIIIGGLIVGYFSKNLKEAVA  
AGLALSIFIFGAFLAYLWGMLEKFLTLSPLYISILLMALAVISATITNFFSPFA  
VKQS"  
5187. .5585  
/gene="AF0489"  
5187. .5585  
/gene="AF0489"  
/note="similar to GB:L77117 SP:Q57917 PID:1591198 percent  
identity: 34.23; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAB90748.1"  
/db\_xref="GI:2650137"  
/translation="MHGSRGNFRNHHKLFTRSKAIVKGLPPLSRSRKRVIAFRIIA  
EKKIDERSLSRALSEKMLSLFGCEFAASGLRLEAFDGERGIVRCYREALDKVMVALTL  
MTHVGGVRVPIPLTLGVSGTIKRCRKRYLEV"  
5588. .6328  
/gene="AF0490"  
5588. .6328  
/gene="AF0490"  
/note="similar to GB:L77117 SP:Q60177 PID:1591299 percent  
identity: 60.82; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="proteasome, subunit alpha (psmA)"

gene

CDS

/protein\_id="AAB90747.1"  
/db\_xref="GI:2650136"  
/translation="MHLPMQGYDRAITVFSPDGRLFQVEYAREAVKRGATAIGIKCKE  
GVILADKRVGSKLLEADTIEKIYKIDEHICAAATSGLVADARVLIDRARIETAOINRLT  
YDEPITVKELAKKICDFKQYTOYGGVRFPGVSLLIAGVDEVPKLYETDPSGALLEYK  
ATAIGMGRNAVTEFFEKEYRDDLSFDDAMVLGLVAMGLSIESELVPENIEVGIVKVVDD  
RTFEVSPSELKPYVERANERIRELLK"  
6332. .7036  
/gene="AF0491"  
6332. .7036  
/gene="AF0491"  
/note="similar to GB:L77117 SP:Q58011 PID:1592297 percent  
identity: 57.33; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAB90746.1"  
/db\_xref="GI:2650135"  
/translation="MVSLLDKAVIARLRKGGEEFVLVDPYLARDLKEGKEVNFEDLLA  
AEEVFKDAKKGERSVDELKRFITGDDVFEIARKIILEGEVQITAEQRREMLEAKRKQ  
IINFISRNITDPRTNAPPPSRIERALEEAKVHIDIFKSVEAQVKDIDVAKLKPILPLK  
FEEMEIAIKIPPEHTGRAISALYNFGGVTRREWQDGSWICVMRIPSGMYGDLMDLLG  
KVAKGEALTKVLRIG"  
Alignment Scores:  
Pred. No.: 247 Length: 12971  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 1 Gaps: 0  
US-09-854-133-587 (1-16) x AE001070 (1-12971)

Oy 6 GlyIleAspPheIleIlePhe 12  
|||||  
Db 9702 GGCATCGATTATATATATTT 9682  
RESULT 33  
AF198256  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
CDS

AF198256  
Haemophilus influenzae genetic island 1, complete sequence.  
AF198256  
AF198256.1 GI:6739645  
Haemophilus influenzae.  
Haemophilus influenzae.  
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.  
1 (bases 1 to 15705)  
Chang,C.C., Gilsdorf,J.R., Dirita,V.J. and Marrs,C.F.  
Identification and genetic characterization of Haemophilus  
influenzae genetic island 1  
Infect. Immun. 68 (5), 2630-2637 (2000)  
20231799  
10768954  
2 (bases 1 to 15705)  
Chang,C.-C., Marrs,C.F., Gilsdorf,J.R. and Dirita,V.J.  
Direct Submission  
Submitted (25-OCT-1999) Epidemiology, University of Michigan, 109  
Observatory Street, Ann Arbor, MI 48109, USA  
Location/Qualifiers  
1. .15705  
/organism="Haemophilus influenzae"  
/strain="Eagan"  
/db\_xref="taxon:727"  
/map="HI0445-HI0446"  
251. .1477  
/note="integrase"  
/codon\_start=1  
/transl\_table=11  
/product="CP4-57 integrase-like protein"

AF198256  
Haemophilus influenzae genetic island 1, complete sequence.  
AF198256  
AF198256.1 GI:6739645  
Haemophilus influenzae.  
Haemophilus influenzae.  
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.  
1 (bases 1 to 15705)  
Chang,C.C., Gilsdorf,J.R., Dirita,V.J. and Marrs,C.F.  
Identification and genetic characterization of Haemophilus  
influenzae genetic island 1  
Infect. Immun. 68 (5), 2630-2637 (2000)  
20231799  
10768954  
2 (bases 1 to 15705)  
Chang,C.-C., Marrs,C.F., Gilsdorf,J.R. and Dirita,V.J.  
Direct Submission  
Submitted (25-OCT-1999) Epidemiology, University of Michigan, 109  
Observatory Street, Ann Arbor, MI 48109, USA  
Location/Qualifiers  
1. .15705  
/organism="Haemophilus influenzae"  
/strain="Eagan"  
/db\_xref="taxon:727"  
/map="HI0445-HI0446"  
251. .1477  
/note="integrase"  
/codon\_start=1  
/transl\_table=11  
/product="CP4-57 integrase-like protein"

AF198256  
Haemophilus influenzae genetic island 1, complete sequence.  
AF198256  
AF198256.1 GI:6739645  
Haemophilus influenzae.  
Haemophilus influenzae.  
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.  
1 (bases 1 to 15705)  
Chang,C.C., Gilsdorf,J.R., Dirita,V.J. and Marrs,C.F.  
Identification and genetic characterization of Haemophilus  
influenzae genetic island 1  
Infect. Immun. 68 (5), 2630-2637 (2000)  
20231799  
10768954  
2 (bases 1 to 15705)  
Chang,C.-C., Marrs,C.F., Gilsdorf,J.R. and Dirita,V.J.  
Direct Submission  
Submitted (25-OCT-1999) Epidemiology, University of Michigan, 109  
Observatory Street, Ann Arbor, MI 48109, USA  
Location/Qualifiers  
1. .15705  
/organism="Haemophilus influenzae"  
/strain="Eagan"  
/db\_xref="taxon:727"  
/map="HI0445-HI0446"  
251. .1477  
/note="integrase"  
/codon\_start=1  
/transl\_table=11  
/product="CP4-57 integrase-like protein"

AF198256  
Haemophilus influenzae genetic island 1, complete sequence.  
AF198256  
AF198256.1 GI:6739645  
Haemophilus influenzae.  
Haemophilus influenzae.  
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.  
1 (bases 1 to 15705)  
Chang,C.C., Gilsdorf,J.R., Dirita,V.J. and Marrs,C.F.  
Identification and genetic characterization of Haemophilus  
influenzae genetic island 1  
Infect. Immun. 68 (5), 2630-2637 (2000)  
20231799  
10768954  
2 (bases 1 to 15705)  
Chang,C.-C., Marrs,C.F., Gilsdorf,J.R. and Dirita,V.J.  
Direct Submission  
Submitted (25-OCT-1999) Epidemiology, University of Michigan, 109  
Observatory Street, Ann Arbor, MI 48109, USA  
Location/Qualifiers  
1. .15705  
/organism="Haemophilus influenzae"  
/strain="Eagan"  
/db\_xref="taxon:727"  
/map="HI0445-HI0446"  
251. .1477  
/note="integrase"  
/codon\_start=1  
/transl\_table=11  
/product="CP4-57 integrase-like protein"

AF198256  
Haemophilus influenzae genetic island 1, complete sequence.  
AF198256  
AF198256.1 GI:6739645  
Haemophilus influenzae.  
Haemophilus influenzae.  
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.  
1 (bases 1 to 15705)  
Chang,C.C., Gilsdorf,J.R., Dirita,V.J. and Marrs,C.F.  
Identification and genetic characterization of Haemophilus  
influenzae genetic island 1  
Infect. Immun. 68 (5), 2630-2637 (2000)  
20231799  
10768954  
2 (bases 1 to 15705)  
Chang,C.-C., Marrs,C.F., Gilsdorf,J.R. and Dirita,V.J.  
Direct Submission  
Submitted (25-OCT-1999) Epidemiology, University of Michigan, 109  
Observatory Street, Ann Arbor, MI 48109, USA  
Location/Qualifiers  
1. .15705  
/organism="Haemophilus influenzae"  
/strain="Eagan"  
/db\_xref="taxon:727"  
/map="HI0445-HI0446"  
251. .1477  
/note="integrase"  
/codon\_start=1  
/transl\_table=11  
/product="CP4-57 integrase-like protein"

AF198256  
Haemophilus influenzae genetic island 1, complete sequence.  
AF198256  
AF198256.1 GI:6739645  
Haemophilus influenzae.  
Haemophilus influenzae.  
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.  
1 (bases 1 to 15705)  
Chang,C.C., Gilsdorf,J.R., Dirita,V.J. and Marrs,C.F.  
Identification and genetic characterization of Haemophilus  
influenzae genetic island 1  
Infect. Immun. 68 (5), 2630-2637 (2000)  
20231799  
10768954  
2 (bases 1 to 15705)  
Chang,C.-C., Marrs,C.F., Gilsdorf,J.R. and Dirita,V.J.  
Direct Submission  
Submitted (25-OCT-1999) Epidemiology, University of Michigan, 109  
Observatory Street, Ann Arbor, MI 48109, USA  
Location/Qualifiers  
1. .15705  
/organism="Haemophilus influenzae"  
/strain="Eagan"  
/db\_xref="taxon:727"  
/map="HI0445-HI0446"  
251. .1477  
/note="integrase"  
/codon\_start=1  
/transl\_table=11  
/product="CP4-57 integrase-like protein"

AF198256  
Haemophilus influenzae genetic island 1, complete sequence.  
AF198256  
AF198256.1 GI:6739645  
Haemophilus influenzae.  
Haemophilus influenzae.  
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.  
1 (bases 1 to 15705)  
Chang,C.C., Gilsdorf,J.R., Dirita,V.J. and Marrs,C.F.  
Identification and genetic characterization of Haemophilus  
influenzae genetic island 1  
Infect. Immun. 68 (5), 2630-2637 (2000)  
20231799  
10768954  
2 (bases 1 to 15705)  
Chang,C.-C., Marrs,C.F., Gilsdorf,J.R. and Dirita,V.J.  
Direct Submission  
Submitted (25-OCT-1999) Epidemiology, University of Michigan, 109  
Observatory Street, Ann Arbor, MI 48109, USA  
Location/Qualifiers  
1. .15705  
/organism="Haemophilus influenzae"  
/strain="Eagan"  
/db\_xref="taxon:727"  
/map="HI0445-HI0446"  
251. .1477  
/note="integrase"  
/codon\_start=1  
/transl\_table=11  
/product="CP4-57 integrase-like protein"



/protein\_id="AAF27347.1"  
/db\_xref="GI:6739646"  
/translation="MARIVKPLSPTEIKNAKPEKEYTILSDGEGLLLLLILPSSGKSWR  
FNYARPLTGKRTKMGALGYPPELSADAKREYRVLLAKGIDPQEEKIRIQOYEYENR  
LKDTFHSVAESYFNGIYKEKAKNPETREKNWELKHIPPYIGDKHVSEIKVELYSI  
YEKIADRSNTLKKIHLQGAIMDAITKGIIESHNCRLAVKNFYIKSSTPHPTIKLDE  
LSKLFQDLENARIGKTKYLLICWSELTALRPKEAVNAEWSIEDFNKLWNIPEKMKG  
QADKKRPHTVPLSSQAIQLEVMKLFSENSESVFAGRSSKNQPMNKATVNVALKHIGY  
KDKLTAHGIRAFIKTFLASHKVERNVSETILSHLLEGGDDLENTYNNRYNLEERKPYM  
OLIGDYCESGMNLNL"  
complement(1649. .3418)  
/note="primase"  
/codon\_start=1  
/transl\_table=11  
/product="phage phi-R73 primase-like protein"  
/protein\_id="AAF27348.1"  
/db\_xref="GI:6739647"  
/translation="MAKLINAPHLADQIHEPYSIDIFILAGRKAWQALDNGKGQEWLLL  
CSLIHGLDRNIKPVILAEKQLENISSIRIVKEDQQSVKLVQYGLAQAEITAICQNL  
KNSDAIDVKLLDAAQTKEDLSAYIQRLRNDKETADLAEQ LAPPEKLEKDGTKKSR  
AFQWLNDMALQRCREIYAYDGKTWNKQENDDLEEKAVKFLDENEFNYSSTIERL  
IKTLKAQLPRMGEMSNDLIAFENGVLNNTMEFESHNRQNLWLTSCIPHKYDKQATDTP  
LFDKWLSPVSGDNKDKARNILAVLYAILTNRYNQWFFEITGKGSKSVFASIAITLL  
AGVKN TASNLEKFDDEGLSLENKTLILCPEQSKYAGDGSGLKSTGGDTVVRYN  
YQDPFDVKITALVMLINRNPCTERSGGVDRRRVIFDEKKIVPEDERDPHFMDKITL  
EVGGIIRKVDSPNDPAKKAQMESEAEVKKLSDPLTDFGFGYFTTEQTDGL  
FIGVTNMGDKIRTHLYPAYLAYTTKAMNIGELGLNNFVIGVEQALKQNGKNHFMKRRH  
TKTGRTNIHFKDFDSFRNEMFS"  
complement(3447. .3797)  
/note="ORF1"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF27349.1"  
/db\_xref="GI:6739648"  
/translation="MSKTAFLVTAGVHVFEQSGFMPIVIAMRESKEAILELHRTYTLQG  
KPTNIRAINLNHYTLEISLGNARAPIIHAEGKRERAEAEAIAPFKTLYPTAIRVIQ  
FKDAFLFKQCAGLD"  
complement(3790. .4428)  
/note="ORF2"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF27350.1"  
/db\_xref="GI:6739649"  
/translation="MKPSNPMQQLQWKKNEQRKINQFVNMNRTKNLASDSGVSGN  
GVHPNKLKIELVKNCSAKQIVKVQNVNLHSEQNTQGNARSEAKSKYQGKLYANPLAF  
KFAQLTRQFKLILDNSRKCLEVYPDDFHKLKMRDEMDMLDIDLKAGKGLFNALAKSQ  
GGTFCRDNQATLKDFNQANGVLIHKFGEVVAQIERLNIERVEGEKLOGASNE"  
complement(4801. .5172)  
/note="ORF3"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF27351.1"  
/db\_xref="GI:6739650"  
/translation="MFNQSILSALQVMADFNINNYKQIPIPMSEKLTDLRSESNL  
CESIEYGLMHLGDLMTQTLGNLADTEQDFTREAMSNDNVKHIGGLIKANAYLLNALRET  
ASLSEYLAGGLDGEIERND"  
complement(5165. .6094)  
/note="ORF4"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF27352.1"  
/db\_xref="GI:6739651"  
/translation="MNSKHSYFIQFTKNFNISALQSAVNAKVFQAQSKTTASREKLN  
YLHQANDSTPENRAFFIRNIPTQSNADVFVNLNPIILSMVERNEQSLTGCLPYAVF  
HLVTFYRPVESLAVVLENKHMELTQMYQIFALIRAPQIKIRLLADTEQQAARSFTD

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

GDTLFVGRINQNPLKNNRTSSIYGNHSLPKCEDAGSIESTTRKGNRNPYQCGIFL  
PKIHSLHVPEKSGALSYIEFAVRLDRNKALSTNQIRSFGRGVVETVSHPMGDTSKNTK  
HGNTPMKIYPKNNRTLATFTPLSVPTQGGANV"  
complement(6261. .6572)  
/note="ORF5"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF27353.1"  
/db\_xref="GI:6739652"  
/translation="MNEARKPTQFLKVLHRLIFSSISGIDGYSMGMTSARNYISELER  
NHLTGKVKRTTEKTADGMGQYRYEIAEAQLKQVIAIYKAKGGELTAHEEQAYFRE  
R"  
complement(6966. .7946)  
/note="ORF6"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF27354.1"  
/db\_xref="GI:6739653"  
/translation="MRKGHFMELLPDYYSLNQSVDFINQRTNSAIKENLLYSVAIEE  
KIRFLQIEIKDQLVKIGRNESEGGFISDDSEIYPKERSIECDKIGLHLKDCSSSL  
SIVHNEITLDGDDLYLSFNKAKITEYKGYIVLHPELLDLFCNQLPQHGKVTNVL  
LDEFNVQTPFEADIFTSFNMEYPKYNEDERLYIKHTFRVNFSDIKISYDDLIKLPF  
SENAREQANLKQETAKLKEILDKAEANNQKQISEQNSPILLGVHRNDDLLKIAI  
EVRNKYWADYPENVKSNQIRDIYIIREYGVARTTAETIEKIACPINRKN"  
complement(8025. .8411)  
/note="ORF7"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF27355.1"  
/db\_xref="GI:6739654"  
/translation="MIKSVLSTFGSFVSALDFLLFLAILLFIIGFLAFIFWPILKWPL  
LAFICAIVFFCYLIYKIKERKPLEQDEVLSDWAKQELQRPITQIRILQKQEQNKPFI  
TGAITHVSDDGKETRLGNITINLKDK"  
complement(8652. .8936)  
/note="ORF8"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF27356.1"  
/db\_xref="GI:6739655"  
/translation="MRCKQAKQNLLSAVNHYKKSTALTFTFVSLYDDEEPAISEVIH  
ALKCKNAAKREIDSRQNSPNMEVLETIYHIAHKNLEDMKKAERRIAKRR"  
complement(8933. .10600)  
/codon\_start=1  
/transl\_table=11  
/product="phage D3 terminase-like protein"  
/protein\_id="AAF27357.1"  
/db\_xref="GI:6739656"  
/translation="MEIWHEYAKKVQSGEIVACRKIKQAVARYFDDLNPAYFFDESA  
VNKFLAFSLCPHVKGHLRGQPIELSDWQTFELFANLLGFKRTDTGLRKYRSAYIQVAR  
KNAKSTVAAVLANWFLLMESGQODIYTAAVSRDQARIYFDDARQMCLLSPPLRKLNI  
QOHLINPKNSILMRPLAAKSSTIEGTNPSLAIVDEYHLHTDNSVYSALELQGARPE  
GLLFAITTAGSNVISACKQHYDYCAQILEGNEQDLSFLVILFELDESEIDNPENWIK  
ANPNIGKSTPYLDFENTIKKARGIPSEWEMLTFRFNWCQGTTPWLGEQNWQAQARN  
YTESDLLHQDCYGLDLSSTNDLTSLCYTFPOGKKVRLITRHYIPEFQLNNVANKNRA  
IYRNWVRQGWLIATEGDCIDYDKIRDDILKDAENFNKIMIGFDVWNATHLRTQLQAAG  
LEVEFPQTYQRFSPVAKSAEVLINRQVIEHNGDPVLSWALSHVVMETDANANIKPNK  
KKAANKIDPAVAFMLMSFGTYQLEYGDLIFELSEEKHALEQFNGIDL"  
complement(10607. .10975)  
/note="ORF9"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF27358.1"  
/db\_xref="GI:6739657"

CDS  
/translation="MTKRTKYKAPDFLDELAKSQWKARITQLSYRGDIKPEDLTNLEI  
YCINYSLFSAVADIAGKNGFSIVNSQGTQSRNPALSAKADAQKVMIKMSALLGDFPVS  
RRKNPVETDVTMLDEILTM"  
complement(11172..11558)  
/note="holin"  
/codon\_start=1  
/transl\_table=11  
/product="phage phi-105 holin-like protein"  
/protein\_id="AAF27359.1"  
/db\_xref="GI:6739658"  
/translation="MPYQPLRRCSYPGCRNVKAGRCCEHKKPKDTRPSSSARGYDHWK  
SKYRAQYLKHPCLVCMLEQKQYTPATVIDHIKPVENGQADPLFWASNHQPLCRDCH  
SYKTRVIDQRFQAKKDGPNPSIIKQ"  
complement(11573..11887)  
/note="ORF10"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF27360.1"  
/db\_xref="GI:6739659"  
/translation="MVEINLTIDEIKAHNLNLDHLDLDELLEAYKSTALEVCRKHIGKT

CDS

Alignment Scores:  
Pred. No.: 293 Length: 15705  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 1 Gaps: 0

CDS

US-09-854-133-587 (1-16) x AF198256 (1-15705)

QY 9 PheIleIlePheTrpIlePhe 15  
|||||  
Db 7380 TTCATTATATTTGGATATTC 7400

RESULT 34

AX251764

LOCUS AX251764 15881 bp DNA linear PAT 05-OCT-2001  
DEFINITION Sequence 25 from Patent WO0168911.  
ACCESSION AX251764  
VERSION AX251764.1 GI:15985119  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial construct.  
REFERENCE 1 (bases 1 to 15881)  
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.  
TITLE Diagnosis of diseases associated with the cell cycle  
JOURNAL Patent: WO 0168911-A 25 20-SEP-2001;  
Epigenomics AG (DE)

FEATURES

source  
Location/Qualifiers  
1..15881  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 4515 a 204 c 3588 g 7574 t  
ORIGIN

Alignment Scores:  
Pred. No.: 296 Length: 15881  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX251764 (1-15881)

QY 10 IleIlePheTrpIlePheTrp 16

|||||  
Db 11862 ATTATTTTGGATATTTGG 11882

RESULT 35

AX344178

LOCUS AX344178 15881 bp DNA linear PAT 01-FEB-2002  
DEFINITION Sequence 25 from Patent WO0200926.  
ACCESSION AX344178  
VERSION AX344178.1 GI:18492066  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 artificial sequences.

AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.  
TITLE Diagnosis of diseases associated with signal transduction  
JOURNAL Patent: WO 0200926-A 25 03-JAN-2002;  
Epigenomics AG (DE)

FEATURES

source  
Location/Qualifiers  
1..15881  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 4515 a 204 c 3588 g 7574 t  
ORIGIN

Alignment Scores:  
Pred. No.: 296 Length: 15881  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX344178 (1-15881)

QY 10 IleIlePheTrpIlePheTrp 16

|||||  
Db 11862 ATTATTTTGGATATTTGG 11882

RESULT 36

AX345162

LOCUS AX345162 15881 bp DNA linear PAT 01-FEB-2002  
DEFINITION Sequence 233 from Patent WO0200928.  
ACCESSION AX345162  
VERSION AX345162.1 GI:18493048  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial sequences.

AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.  
TITLE Diagnosis of diseases associated with the immune system  
JOURNAL Patent: WO 0200928-A 233 03-JAN-2002;  
Epigenomics AG (DE)

FEATURES

source  
Location/Qualifiers  
1..15881  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 4515 a 204 c 3588 g 7574 t  
ORIGIN

Alignment Scores:  
Pred. No.: 296 Length: 15881  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX345162 (1-15881)

QY 10 IleIlePheTrpIlePheTrp 16

|||||  
Db 11862 ATTATTTTGGATATTTGG 11882

Db 11862 ATTATTTTGGATATTTGG 11882

RESULT 37  
AX348575

LOCUS AX348575 15881 bp DNA linear PAT 06-FEB-2002  
DEFINITION Sequence 33 from Patent WO0202807.  
ACCESSION AX348575  
VERSION AX348575.1 GI:18614610  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequences.

REFERENCE 1  
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.  
TITLE Diagnosis of diseases associated with cell signalling  
JOURNAL Patent: WO 0202807-A 33 10-JAN-2002;  
Epigenomics AG (DE)

FEATURES  
source  
1. .15881  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 4515 a 204 c 3588 g 7574 t  
ORIGIN

Alignment Scores:  
Pred. No.: 296 Length: 15881  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX348575 (1-15881)

QY 10 lleilepHeTrpIlePhe 16  
|||||

Db 11862 ATTATTTTGGATATTTGG 11882

RESULT 38  
AC017868/c

LOCUS AC017868 16023 bp DNA linear HTG 09-DEC-1999  
DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
ACCESSION AC017868  
VERSION AC017868.1 GI:65533322  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 16023)  
Adams, M. and Venter, J.C.  
Direct Submission  
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
This sequence was identified as CDM:10212420 by the submitter.  
For more information on this record e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
source  
1. .16023  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"

BASE COUNT 5472 a 3032 c 2604 g 4915 t  
ORIGIN

Alignment Scores:  
Pred. No.: 299 Length: 16023  
Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 2 Gaps: 0

US-09-854-133-587 (1-16) x AC017868 (1-16023)

QY 9 PheileilepHeTrpIlePhe 15  
|||||

Db 2500 TTCATAATTTTGGATATTT 2480

RESULT 39  
AX344576

LOCUS AX344576 16633 bp DNA linear PAT 01-FEB-2002  
DEFINITION Sequence 1 from Patent WO0200927.  
ACCESSION AX344576  
VERSION AX344576.1 GI:18492462  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequences.

REFERENCE 1  
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.  
TITLE Diagnosis of diseases associated with development genes  
JOURNAL Patent: WO 0200927-A 1 03-JAN-2002;  
Epigenomics AG (DE)

FEATURES  
Location/Qualifiers  
source  
1. .16633  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 4452 a 142 c 3667 g 8372 t  
ORIGIN

Alignment Scores:  
Pred. No.: 309 Length: 16633  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX344576 (1-16633)

QY 9 PheileilepHeTrpIlePhe 15  
|||||

Db 11972 TTTATTTTGGATTTT 11992

RESULT 40  
U42843/c

LOCUS U42843 23933 bp DNA linear INV 04-JUL-2002  
DEFINITION Caenorhabditis elegans cosmid T19E7, complete sequence.  
ACCESSION U42843  
VERSION U42843.2 GI:14574233  
KEYWORDS HTG.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 23933)  
Waterston, R.  
Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
Science 282 (5396), 2012-2018 (1998)  
99069613  
MEDLINE  
PUBMED 9851916  
REFERENCE 2 (bases 1 to 23933)  
AUTHORS Pauley, A. and Gattung, S.  
TITLE The sequence of C. elegans cosmid T19E7  
JOURNAL Unpublished (2001)  
REFERENCE 3 (bases 1 to 23933)  
AUTHORS Waterston, R.  
TITLE Direct Submission

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Submitted (10-DEC-1995) Robert Waterston  
4 (bases 1 to 23933)  
Waterston,R.  
Direct Submission  
Submitted (28-JUN-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
5 (bases 1 to 23933)  
Waterston,R.  
Direct Submission  
Submitted (23-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
6 (bases 1 to 23933)  
Waterston,R.  
Direct Submission

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Submitted (04-JUL-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 28, 2001 this sequence version replaced gi:1125791.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: rw@nematode.wustl.edu and jesus@sanger.ac.uk

COMMENT

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all  
regions were double stranded, sequenced with an alternate chemistry  
or covered by high quality data (i.e., phred quality >= 30); an  
attempt was made to resolve all sequencing problems, such as  
compressions and repeats; all regions were covered by sequence from  
more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its  
analysis see:

<http://www.wormbase.org/db/seq/sequence?name=Tl9E7;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is B0547, 1400 bp overlap; the 3' cosmid is C55C3,  
200 bp overlap. Actual start of this cosmid is at base position 401  
of Tl9E7; actual end is at 12238 of C55C3.

NOTES:

Coding sequences below are the result of integration and manual  
review of the following data : computer analysis using the program  
Genefinder (P. Green and L. Hillier, personal communication), the  
large scale EST projects of Yuji Kohara  
([http://www.ddbj.nig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html)) and The C.  
elegans ORFeome cloning project (<http://worfdb.dfci.harvard.edu/>),  
similarity to other proteins from BlastX analyses  
(<http://blast.wustl.edu/>), sequence conservation with C. briggsae  
using Jim Kent's WABA alignment program (Genome Research  
10:1115-1125, 2000), individual C. elegans GenBank submissions,  
and personal communications with C. elegans researchers. tRNAs  
are predicted using the program tRNAscan-SE (Lowe, T.M. and  
Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES  
source

Location/Qualifiers  
1. .23933  
/organism="Caenorhabditis elegans"  
/strain="Bristol N2"  
/db\_xref="taxon:6239"  
/chromosome="IV"  
/clone="Tl9E7"  
complement(36..8656)

gene

/gene="skn-1"  
/note="for a graphical representation of this gene see:  
<http://www.wormbase.org/db/seq/sequence?name=Tl9E7.2a;class=Sequence>"  
complement(join(36..112,165..532,686..944,1621..1773,  
1827..2148,3455..3879,4042..4180,8528..8656))  
/gene="skn-1"  
/note="coded for by the following C. elegans cDNAs:  
yk287d4.3, yk18e11.5, yk155h9.5, CESAB55F, yk194d1.5,  
yk202g2.5, yk250e4.5, yk287d4.5, yk314h8.5, yk320b11.5,  
yk385h2.3, yk403d12.5, yk415e5.5, yk421f9.5, yk204f2.5,  
yk251g1.5, yk291h2.5, yk311c1.5, yk2d12.3, CEMSE62F,  
CESAC19F, yk250e4.3, yk415e5.3, yk251g1.3, yk291h2.3"  
/codon\_start=1  
/product="C. elegans SKN-1 protein (corresponding sequence  
Tl9E7.2a)"  
/protein\_id="AAA83594.2"  
/db\_xref="GI:14574234"  
/translation="MGGSSRRQSTSATRRDQKRRRRQCFSSVADDEETTSIYGVSS  
IFIWILATSSLILVISSPSNTSIQSSYDRITTKHLLDNISPTFKMYTDSNNRNFDE  
VNHQHQEQDENGQSKYDYPQFNRPMLRWDDQRMMEYFMSNGPVEVTPVMPILTEH  
PPASFGRGSPSTERPTTSRYEYSSPSLEDIDLIDVLWRSIDIAEGKTRQVAPADQYE  
CDLQTLTEKSTVAPLTAENARYEDLSKGYNGFFESFNNNQYQKHQQQREQIKTP  
TLEHPTQKAELEDLDFEDLAQLFEDVSREEGQNLQFNKQHPVINNVSLSEGIY  
NQANLTEMQEMRDCNQVSIPTTSTAQETLENVTDSTQVQWLPTVEVPNDVFP  
TSNYAYIGMNDSLQAVVNGQIDYDHSYQSTGTPLSPLIGSSGRQQQTQTSPGSV  
TVTATATQSLFDPYHSQRHSFSDCTDSSSTCSRLSSSRKQKRSKDEQLASDNELPVSFAFQ  
LAPSSGSRYSRQSSSSQSSIKIARVVPLASQGRKGRQSKDEQLASDNELPVSFAFQ  
ISEMSLSELQVLKNESEYQRLIRKIRRRGKNKVAARTCRQRTDRHDKMSHYI"  
complement(join(36..112,165..532,686..944,1621..1773,  
1827..2148,3455..3877))  
/gene="skn-1"  
/note="coded for by the following C. elegans cDNAs:  
yk782b12.5, yk782b12.3"  
/codon\_start=1  
/product="C. elegans SKN-1 protein (corresponding sequence  
Tl9E7.2c)"  
/protein\_id="AAM75366.1"  
/db\_xref="GI:21693910"  
/translation="MYTDSNNRNFDEVNHQHQEQDENGQSKYDYPQFNRPMLRWDRD  
QRMMEYFMSNGPVEVTPVMPILTEHPPASFGRGSPSTERPTTSRYEYSSPSLEDID  
LIDVLWRSIDIAEGKTRQVAPADQYECDLQTLTEKSTVAPLTAENARYEDLSKGFYN  
GFFSFNNNQYQKHQQQREQIKTPLEHPTQKAELEDLDFEDLAQLFEDVSREEG  
QNLQFNKQHPVINNVSLSEGIYVQANLTEMQEMRDCNQVSIPTTSTAQPE  
TLFNVTDSTQVQWLPTVEVPNDVFPVTSNYAYIGMNDSLQAVVNGQIDYDHSYQST  
GQTPLSPLIGSSGRQQQTQTSPGSVTVTATATQSLFDPYHSQRHSFSDCTDSSSTC  
SRLSSSPRYTSESSTGTHESRFYKGLAPSSGRYSRQSSSSQSSIKIARVVPLAS  
GQKRGKRSKDEQLASDNELPVSFAFQISEMSLSELQVLKNESEYQRLIRKIRRR  
GKNKVAARTCRQRTDRHDKMSHYI"  
complement(join(36..112,165..532,686..944,1139..1367))  
/gene="skn-1"  
/note="coded for by the following C. elegans cDNAs:  
yk774f09.5, yk774f09.3"  
/codon\_start=1  
/product="C. elegans SKN-1 protein (corresponding sequence  
Tl9E7.2b)"  
/protein\_id="AAM75365.1"  
/db\_xref="GI:21693909"  
/translation="MSLPDFASSLLASSTTTNTTAPAAVNSFDEQEESKKILNM  
YLQMFNQVQDQGHGHHQHPYAYSGVSTFDVFPTSNYAYIGMNDSLQAVVNGQI  
DYDHSYQSTGTPLSPLIGSSGRQQQTQTSPGSVTVTATATQSLFDPYHSQRHSFSD  
CTDSSSTCSRLSSSPRYTSESSTGTHESRFYKGLAPSSGRYSRQSSSSQSSIK  
IARVVPLASGQKRGKRSKDEQLASDNELPVSFAFQISEMSLSELQVLKNESEYQRL  
QIRKIRRRGKNKVAARTCRQRTDRHDKMSHYI"  
2259..3224  
/gene="Tl9E7.1"  
/note="for a graphical representation of this gene see:  
<http://www.wormbase.org/db/seq/sequence?name=Tl9E7.1;class=Sequence>"  
join(2259..2414,2800..2950,2995..3224)  
/gene="Tl9E7.1"  
/note="contains similarity to Pfam domain PF00059  
(lectin\_c), Score=82.4, E-value=2.9e-21, N=1; coded for by

CDS

gene

CDS

CDS

CDS



the following C. elegans cDNAs: CEESH64R, CEMSH13F"

/codon\_start=1  
/product="Hypothetical protein T19E7.1"  
/protein\_id="AAA83595.1"  
/db\_xref="GI:1125793"  
/translation="MILLLLYLLIPTINTVWIPSLKSSYSQSTAGQRAINLDVNVRL  
TVLALQSGWQLYEDHWYKMFETVMWIPAEVNCVCRSMGGHLVSKDESENLFVHKLRK  
KNNIWIQGLKLNDFHVKYKWSGSEADYLNWASSQNPDPDCAYNMAFHQEQRTWFD  
YGCREMPLQFFVCKLPMP"  
complement(9612. .12689)  
/gene="T19E7.3"  
/note="for a graphical representation of this gene see:  
http://www.wormbase.org/db/seq/sequence?name=T19E7.3;class  
=Sequence"  
complement(join(9612. .9797,10409. .10618,10813. .10969,  
11016. .11090,11637. .11938,11985. .12083,12591. .12689))  
/gene="T19E7.3"  
/note="coded for by the following C. elegans cDNAs:  
yk391b8.5, yk486f7.5, yk610c10.5, yk656b6.5"  
/codon\_start=1  
/product="Hypothetical protein T19E7.3"  
/protein\_id="AAA83597.2"  
/db\_xref="GI:14574235"  
/translation="MTQRSHICLNCQHPLRLDFTQRRPDSADSEKKSEVITEALTG  
HSRNLMLKLSIDAPPSDAPVNCDSALRNEMDAQVATLDDEIKTYQTYINYLKENHP  
TTSIPDLKAKLQNVSDSEKELEQQLKLLAEEDLDLQTKRRTAEAASEKSGELWK  
KYRNLNRQVFEDQDLHSLAEAEQYAEVQHRKLDTDNLDFCFHIWVDGIVGEINGFR  
LGYLKDAPVEFTEINAALQIVLLLEILLERIGVQHHELMVPMGSHSYIKLRNGID  
METYALYQGTPLSGSGIDPQIRRFLLQLLEFLKELKDRNKNEKPPYQIHADSLVDN  
GVKYNVMTLNTDVRWTRAMALMTDLKAACAQCDALRSPI"  
complement(20905. .22799)

gene

Alignment Scores:

Pred. No.:	427	Length:	23933
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	43.75%	Indels:	0
DB:	3	Gaps:	0

US-09-854-133-587 (1-16) x U42843 (1-23933)

QY 7 IleAspPheIleIlePheTrp 13  
|||||

Db 9853 ATCGATTTTATAATTTTGG 9833

RESULT 41

U42841

LOCUS U42841 31228 bp DNA linear INV 23-MAY-2002

DEFINITION Caenorhabditis elegans cosmid T17H7, complete sequence.

ACCESSION U42841

VERSION U42841.1 GI:2072756

KEYWORDS HTG.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 31228)

AUTHORS Waterston,R.

TITLE Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
Science 282 (5396), 2012-2018 (1998)

JOURNAL 99069613

MEDLINE 9851916

PUBMED 9851916

REFERENCE 2 (bases 1 to 31228)

AUTHORS Favello,A.

TITLE The sequence of C. elegans cosmid T17H7

JOURNAL Unpublished (2001)

REFERENCE 3 (bases 1 to 31228)

AUTHORS Waterston,R.

TITLE Direct Submission

JOURNAL Submitted (25-JUL-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

REFERENCE

AUTHORS Louis, MO 63110, USA

TITLE 4 (bases 1 to 31228)

JOURNAL Waterston,R.

Direct Submission

Submitted (14-SEP-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

5 (bases 1 to 31228)

Waterston,R.

Direct Submission

Submitted (23-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

On May 7, 1997 this sequence version replaced gi:1125779.

Submitted by:

Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis , MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: rw@nematode.wustl.edu and jesusanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all  
regions were double stranded, sequenced with an alternate chemistry  
or covered by high quality data (i.e., phred quality >= 30); an  
attempt was made to resolve all sequencing problems, such as  
compressions and repeats; all regions were covered by sequence from  
more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its  
analysis see:  
http://www.wormbase.org/db/seq/sequence?name=T17H7;class=Sequence

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C24A1, 3897 bp overlap; the 3' cosmid is F42G9,  
3000 bp overlap. Actual start of this cosmid is at base position  
3698 of T17H7; actual end is at 19222 of F42G9.

NOTES:

Coding sequences below are the result of integration and manual  
review of the following data : computer analysis using the program  
Genefinder (P. Green and L. Hillier, personal communication), the  
large scale EST projects of Yuji Kohara  
(http://www.ddbj.nig.ac.jp/c-elegans/html/CE\_INDEX.html) and The C.  
elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/),  
similarity to other proteins from BlastX analyses  
(http://blast.wustl.edu/), sequence conservation with C. briggsae  
using Jim Kent's WABA alignment program (Genome Research  
10:115-1125, 2000), individual C. elegans GenBank submissions,  
and personal communications with C. elegans researchers. tRNAs  
are predicted using the program tRNAscan-SE (Lowe, T.M. and  
Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES

source

1. .31228  
/organism="Caenorhabditis elegans"  
/strain="Bristol N2"  
/db\_xref="taxon:6239"  
/chromosome="III"  
/clone="T17H7"  
complement(52. .18951)  
/gene="T17H7.4"  
/note="for a graphical representation of this gene see:  
http://www.wormbase.org/db/seq/sequence?name=T17H7.4a;clas  
s=Sequence"  
complement(join(52. .118,361. .477,530. .839,1295. .1457,  
1506. .1606,1993. .2098,2399. .2557,2722. .2850,4319. .4339,

gene

CDS

11740. .11760,12073. .12171,12221. .12319,12368. .12502,  
15398. .15595,18865. .18951))  
/gene="T17H7.4"  
/note="coded for by the following C. elegans cDNAs:  
yk478b10.5"  
/codon\_start=1  
/product="Hypothetical protein T17H7.4b"  
/protein\_id="AAK77639.1"  
/db\_xref="GI:15011818"  
/translation="MTSHIATETSVNRWSTEPVVREGRPPGPEFPSTKQGREQDGN  
VIGITPNYKIGEOKPNSLNLGKTSGLPRGSDYKMTYRQEDYKQTIIDPKVITNANM  
PYISTRIPDRNLGDFPSEHSKSIKLDPLEQRFRSSSRPITTSNYLAESVDRH  
REMEASRLREYLKAKENANQPNKPGWPKKNDSESLRELETKQRIETLKRDKANG  
PVTVPPLRTKTPPPPPVRRHEIYEQHKRYTSAPNLSAVIRIQDDKPRSIMKRN  
VRREQMLYPTVDTQVKSVMKPTVTETVQFEETRRTEVEVRRVQRREKERRRHH  
SSSRHSGWEGTGGYQHRSSSLSRGGGGGGQETTYRQETRRQQHNNYDDNFNRG  
IAHARYGSLSDSLRGELQYVPNGEVRSFYRGGGQDYRREDAFYRPVSKRSYADWDDAG  
RAGMGREVRYYDDMSRLAEAFRDSLMPMPAGNMNERDHRTEQLPGGYETFNKERHA  
NSGRRSGRDKGKVPDNEATQYNYKREQTLNDDRRR"  
complement(join(52. .118,361. .477,530. .839,1295. .1457,  
1506. .1606,1993. .2098,2399. .2557,3005. .3097,3945. .4031,  
4089. .4217,4319. .4339,11740. .11760,12073. .12171,  
12221. .12319,12368. .12502,15398. .15595,18865. .18951))  
/gene="T17H7.4"  
/note="coded for by the following C. elegans cDNAs:  
yk287e3.3, yk48a6.3, yk366c8.3, yk15d2.3, yk64e10.3,  
yk64e10.5, yk149f7.5, yk366c8.5, yk410e10.3, yk410e10.5,  
yk15d2.5, yk34e11.5, yk287e3.5, yk360d3.5, yk760c12.5,  
yk758c04.5"  
/codon\_start=1  
/product="Hypothetical protein T17H7.4a"  
/protein\_id="AAK77638.1"  
/db\_xref="GI:15011817"  
/translation="MTSHIATETSVNRWSTEPVVREGRPPGPEFPSTKQGREQDGN  
VIGITPNYKIGEOKPNSLNLGKTSGLPRGSDYKMTYRQEDYKQTIIDPKVITNANM  
PYISTRIPDRNLGDFPSEHSKSIKLDPLEQRFRSSSRPITTSNYLAESVDRH  
REMEASRLREYLKAKENANQPNKPGWPKKNDSESLRELETKQRIETLKRDKANG  
PVNAELDELSKRAEELKRDGWSKYKLVEDSIYKTDPPMPANIKDQVRDLLESNSV  
ETTTTQRDHDKSGYVTDVSTATWFSYDYSRVSVMNGASHDILKDDKPRSIMKRN  
DLVREQMLYPTVDTQVKSVMKPTVTETVQFEETRRTEVEVRRVQRREKERRR  
HSSSRHSGWEGTGGYQHRSSSLSRGGGGGGQETTYRQETRRQQHNNYDDNFNRG  
RGIAHARYGSLSDSLRGELQYVPNGEVRSFYRGGGQDYRREDAFYRPVSKRSYADWDD  
DDRRSVSFFRRGQQQVSPFVEFPPTLPRRGGGGDYRREDAFYRPVSKRSYADWDD  
AGRAGMGREVRYYDDMSRLAEAFRDSLMPMPAGNMNERDHRTEQLPGGYETFNKER  
HANSRRSGRDKGKVPDNEATQYNYKREQTLNDDRRR"  
complement(join(52. .118,361. .477,530. .839,1295. .1457,  
1506. .1606,1993. .2098,2399. .2557,2722. .2850,3005. .3097,  
3945. .4031,4089. .4217,4319. .4339,11740. .11760,  
12073. .12171,12221. .12319,12368. .12502,15398. .15595,  
18865. .18951))  
/gene="T17H7.4"  
/note="coded for by the following C. elegans cDNAs:  
yk89d3.3, yk89d3.5"  
/codon\_start=1  
/product="Hypothetical protein T17H7.4c"  
/protein\_id="AAK77640.1"  
/db\_xref="GI:15011819"  
/translation="MTSHIATETSVNRWSTEPVVREGRPPGPEFPSTKQGREQDGN  
VIGITPNYKIGEOKPNSLNLGKTSGLPRGSDYKMTYRQEDYKQTIIDPKVITNANM  
PYISTRIPDRNLGDFPSEHSKSIKLDPLEQRFRSSSRPITTSNYLAESVDRH  
REMEASRLREYLKAKENANQPNKPGWPKKNDSESLRELETKQRIETLKRDKANG  
PVNAELDELSKRAEELKRDGWSKYKLVEDSIYKTDPPMPANIKDQVRDLLESNSV  
ETTTTQRDHDKSGYVTDVSTATWFSYDYSRVSVMNGASHDILKDDKPRSIMKRN  
PPPPVRRHEIYEQHKRYTSAPNLSAVIRIQDDKPRSIMKRNLDLVRREQMLYPTVDT  
QVKSVMKPTVTETVQFEETRRTEVEVRRVQRREKERRRHHSSSRHSGWEGT  
GGYQHRSSSLSRGGGGGGQETTYRQETRRQQHNNYDDNFNRGIAHARYGSLSDSL  
RGELQYVPNGEVRSFYRGGGQDYRREDAFYRPVSKRSYADWDDAGRAGMGREVRYYDD  
DMSRLAEAFRDSLMPMPAGNMNERDHRTEQLPGGYETFNKERHANSRRSGRDKGPV  
DFNEATQYNYKREQTLNDDRRR"  
complement(join(52. .118,361. .477,530. .839,1295. .1457,  
1506. .1606,1993. .2098,2399. .2557,2722. .2850,3005. .3097,

CDS

CDS

CDS

3945. .4031,4089. .4217,4319. .4339,5703. .5801,5848. .7713,  
7771. .8040,11242. .11391))  
/gene="T17H7.4"  
/note="coded for by the following C. elegans cDNAs:  
yk892g06.3, yk892g06.5"  
/codon\_start=1  
/product="Hypothetical protein T17H7.4d"  
/protein\_id="AAC48171.2"  
/db\_xref="GI:15617855"  
/translation="MSFQQQASPISPGDSGLIKMRKNRYLDHTSKVNEPKKYSRMD  
QLPTGKERRASYHVESYIHSQQIQHPQQHPPPPPPVVAETIQHPHPQQQQP  
STTTISSRLYKSLQLLYPPGSTENIASHKPAIRPEKSVYENNPISLSELNVQVKRA  
PQGGPPPPPPQAPADIDYPERLMSPPTHQPRSIKGYSSQYYSQTNLTESHHQH  
SQETSLNPQQMHPRPSSDRRTPTHFEDLSEDEKTRIMHENLQRHNRMRSGPGRPTTT  
SFNGPFERLEQVNPQNPQQPQYQQHPRSQSVDPGDMNGGPPPIHQNFSAEIELH  
HYSRNAEPSVVVPPVSEKERPSVLAKNFNDPERIDEYQQRLEIYEAIQRTNEK  
QAVSMKQVNMAMQQQKLYERSHGMTSPVIMESISPPQVSHHHHQPQQHQHQPPQ  
QYYPSPPPPPQYRDPPEFVNPVQVYETRPISALSDQIDQPPSQAPTSWKRTYIYER  
PRDVAKNEIITSDQLEKEAYDVLKRRRETVEKPEEAPRINRLGRRWQPPPERPYV  
WPTLRAMSVPEPSPVDFAPGVPHNYDDNEEYKWEVVDGPKKEDKNETPVPNSPP  
ASPRRGHGVGLDEPARRQAKYVIQPSDPGSHRPKAVFRKERHAPSGGFYPHAPNAIK  
VVKRAQSVQGLLSPDNDVEVIHQRYHRLDLEQNGHGHQKLRSSQHGSEMDLRRST  
QEMPDEKIEYELPHSSQIVQKMPRHVDVQRRLSKFEGSIQNLRAATSTQHLSDSAN  
KGPVNAELDELSKRAEELKRDGWSKYKLVEDSIYKTDPPMPANIKDQVRDLLESRN  
SVETTTTQRDHDKSGYVTDVSTATWFSYDYSRVSVMNGASHDILKTVPLRTKT  
PPPPVRRHEIYEQHKRYTSAPNLSAVIRIQDDKPRSIMKRNLDLVRREQMLYPTV  
DTQVKSVMKPTVTETVQFEETRRTEVEVRRVQRREKERRSRHSSSRHSGWEG  
HTGGYQHRSSSLSRGGGGGGQETTYRQETRRQQHNNYDDNFNRGIAHARYGSLSD  
SLRRGELQYVPNGEVRSFYRGGGQDYRREDAFYRPVSKRSYADWDDAGRAGMGREVRYYDD  
DMSRLAEAFRDSLMPMPAGNMNERDHRTEQLPGGYETFNKERHANSRRSGRDKGPV  
DFNEATQYNYKREQTLNDDRRR"  
complement(join(52. .118,361. .477,530. .839,1295. .1457,  
1506. .1606,1993. .2098,2399. .2557,2722. .2850,3005. .3097,  
3945. .4031,4089. .4217,4319. .4339,11740. .11760,  
12073. .12171,12221. .12319,12368. .12502,15398. .15595,  
18865. .18951))  
/gene="T17H7.4"  
/note="coded for by the following C. elegans cDNAs:  
yk89d3.3, yk89d3.5"  
/codon\_start=1  
/product="Hypothetical protein T17H7.4e"  
/protein\_id="AAK77641.1"  
/db\_xref="GI:15011820"  
/translation="MTSHIATETSVNRWSTEPVVREGRPPGPEFPSTKQGREQDGN  
VIGITPNYKIGEOKPNSLNLGKTSGLPRGSDYKMTYRQEDYKQTIIDPKVITNANM  
PYISTRIPDRNLGDFPSEHSKSIKLDPLEQRFRSSSRPITTSNYLAESVDRH  
REMEASRLREYLKAKENANQPNKPGWPKKNDSESLRELETKQRIETLKRDKANG  
PVNAELDELSKRAEELKRDGWSKYKLVEDSIYKTDPPMPANIKDQVRDLLESNSV  
ETTTTQRDHDKSGYVTDVSTATWFSYDYSRVSVMNGASHDILKDDKPRSIMKRN  
DLVREQMLYPTVDTQVKSVMKPTVTETVQFEETRRTEVEVRRVQRREKERRR  
HSSSRHSGWEGTGGYQHRSSSLSRGGGGGGQETTYRQETRRQQHNNYDDNFNRG  
RGIAHARYGSLSDSLRGELQYVPNGEVRSFYRGGGQDYRREDAFYRPVSKRSYADWDD  
DDRRSVSFFRRGQQQVSPFVEFPPTLPRRGGGGDYRREDAFYRPVSKRSYADWDD  
AGRAGMGREVRYYDDMSRLAEAFRDSLMPMPAGNMNERDHRTEQLPGGYETFNKER  
HANSRRSGRDKGKVPDNEATQYNYKREQTLNDDRRR"  
complement(join(52. .118,361. .477,530. .839,1295. .1457,  
1506. .1606,1993. .2098,2399. .2557,2722. .2850,3005. .3097,  
3945. .4031,4089. .4217,4319. .4339,11740. .11760,  
12073. .12171,12221. .12319,12368. .12502,15398. .15595,  
18865. .18951))  
/gene="T17H7.4"  
/note="coded for by the following C. elegans cDNAs:  
yk89d3.3, yk89d3.5"  
/codon\_start=1  
/product="Hypothetical protein T17H7.4f"  
/protein\_id="AAK77642.1"  
/db\_xref="GI:15011821"  
/translation="MTSHIATETSVNRWSTEPVVREGRPPGPEFPSTKQGREQDGN  
VIGITPNYKIGEOKPNSLNLGKTSGLPRGSDYKMTYRQEDYKQTIIDPKVITNANM  
PYISTRIPDRNLGDFPSEHSKSIKLDPLEQRFRSSSRPITTSNYLAESVDRH  
REMEASRLREYLKAKENANQPNKPGWPKKNDSESLRELETKQRIETLKRDKANG  
PVNAELDELSKRAEELKRDGWSKYKLVEDSIYKTDPPMPANIKDQVRDLLESNSV  
ETTTTQRDHDKSGYVTDVSTATWFSYDYSRVSVMNGASHDILKDDKPRSIMKRN  
DLVREQMLYPTVDTQVKSVMKPTVTETVQFEETRRTEVEVRRVQRREKERRR  
HSSSRHSGWEGTGGYQHRSSSLSRGGGGGGQETTYRQETRRQQHNNYDDNFNRG  
RGIAHARYGSLSDSLRGELQYVPNGEVRSFYRGGGQDYRREDAFYRPVSKRSYADWDD  
DDRRSVSFFRRGQQQVSPFVEFPPTLPRRGGGGDYRREDAFYRPVSKRSYADWDD  
AGRAGMGREVRYYDDMSRLAEAFRDSLMPMPAGNMNERDHRTEQLPGGYETFNKER  
HANSRRSGRDKGKVPDNEATQYNYKREQTLNDDRRR"  
complement(join(52. .118,361. .477,530. .839,1295. .1457,  
1506. .1606,1993. .2098,2399. .2557,2722. .2850,3005. .3097,

Alignment Scores:

Pred. No.: 541 Length: 31228  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 3 Gaps: 0

US-09-854-133-587 (1-16) x U42841 (1-31228)

QY 7 IleAspPheIleIlePheTrp 13  
|||||

Db 27329 ATTGATTTATTATTATTTTGG 27349

RESULT 42

AF391289 34544 bp DNA linear INV 07-MAY-2002  
Branchiostoma floridae cosmid MPMGcl17B2065, partial sequence.  
AF391289  
AF391289.1 GI:20302742

SOURCE

ORGANISM

Branchiostoma floridae.  
Branchiostoma floridae  
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
Branchiostoma.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

Location/Qualifiers  
1. .34544  
/organism="Branchiostoma floridae"  
/db\_xref="taxon:7739"  
/clone="cosmid MPMGcl17B2065"  
/clone\_lib="Resource Center of the German Human Genome  
Project (RZPD) Berlin, Library n. 117"

mRNA

```
join(<400. .535,951. .1036,1480. .1590,2259. .2449,2909. .3061,
3729. .4273,4966. .5050,5682. .>5781)
/product="unknown"
join(<400. .535,951. .1036,1480. .1590,2259. .2449,2909. .3061,
3729. .4273,4966. .5050,5682. .5781)
/note="CDS1"
```

CDS

```
/codon_start=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAM18870.1"
/db_xref="GI:20302743"
/translation="KAKRAQKGRKRLPGVETEVLRIEIQEEQRREKIRQAKRRSA
GSSQVLDPNTSTTSEFTTASEKDLSDVEGSPSTAEESEATEIPVDIPVVLKDRLEED
HIMICKRGKLVQLPCQPNVITILENYLFFFAAKGTSVDRHLHHQLSRAAQPFAAKFPT
PSNVHSTRVPPPERNVDLVKEVVDGVRIMFDLILPTLLYQDEEVQHTQMATSTFIPL
TPAGQTEADQAATPPPPKIRVTRSASRSPICHEAQKSHPEYPLRDSAVKFSFGQALA
SKSDGLSLRRSRFSDQEPALSLPPTQKRHRSSISHAQPLPAAVRQAEQTPTSPPP
LIPAPPPPGEPATAMDADSKLSSVDTAAEVKVQHKLALLDEAMCKLLPDDFSDQGQN
PPSLIYGAQHLLRLFKVLPPELLGKMLHLPVSKILCYHLQLFLKYLADHRDILLPATA
YVSAAEVLGNLRQPKVIR"
complement(join(<6992. .7192,7693. .7832,8188. .8399,
8980. .9725,10309. .10443,10831. .10943,11221. .>11242))
/product="unknown"
complement(join(6992. .7192,7693. .7832,8188. .8399,
8980. .9725,10309. .10443,10831. .10943,11221. .11242))
/note="CDS2"
```

mRNA

```
/codon_start=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAM18871.1"
/db_xref="GI:20302744"
/translation="MAKRVLAGQADCPYAKVELLADSLQSKLQDFHVKIVKTPPEW
KPWLQQLCADNGWTHGRSPILWRELVDGKGLLIGHNDFMETARYSYGMSGLSE
MMKNIAAENLAVTQELREEEETIRROSKPLHVCVINAARSPAYHVLPVLVNGKILREE
EIALHLHDSEENLEKLGLEMEVFDLSFPFLKEISVTDLPTAFQNAHIAIIVDDFDQ
GKEDAIGDMETKVSFYKRVAAEAINQASKDIRLVAGTGPLNSLVSLIDHTPSPR
QNIAAQAQKQAKSLAKRLTVNSAGVCDVIVGNGVGGTTYTDVSRARVHGDGAI
WGPPSYSCSVSEMVHDNKLWEGEFLEQLQSRSHITQDSLQHSADLSMAAAISSLSYV
WNGSPGQLFLSVCSEGFYNIPEGVVFSPFMFKGSEVVDIDMNEDMRVMLSSI
TAQLVEEKDLLLHPTNANHEKLLKVFVGTDTISTPSSVKLKDIMEETEKSSSEDTPEA
AAAAVSTGDETGPS"
```

CDS

```
/codon_start=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAM18872.1"
/db_xref="GI:20302745"
/translation="MASKTSTWNTKADLVPLEKDKAKELRARISKIKQAEVLRPGAVF
VGHIPHGFEPQVKRFFSQFGTVSRVRLARSKKTGGPKGYGIEFECEEVAKVVAETM
NNYLMFEKLLKCEFIPTERLHPKTFHGSNRKFLKPTRRSTAIRKINKPKDAAATLQSA
MRLASKNLGRKRYAALGLNYDFSGYDDMAKAKLQAAASIVTPKPKSEEVKTTKQE
PATLLDSDPPEISFTPTPPGVSKVKIKKAACKQT"
join(<17025. .17173,17557. .17694,18464. .18964,
19413. .>19716)
/product="unknown"
join(17025. .17173,17557. .17694,18464. .18964,19413. .19716)
/note="CDS4"
```

mRNA

```
/codon_start=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAM18873.1"
/db_xref="GI:20302746"
/translation="MASETAPVWPQNYTSPDKRKEGGIRHRRLLVFLTLLCLSVITNA
LLWLSRQGLTNMLLARRDKNTDGNLSRPVPEVSEAPFSVEDVELRIRNAASKL
HPDRKFRLGKWPVCLRTGKTKSKKHCPPSKHIGNYRCAVGVNGSGLLGSDCGAD
IDAKDYVIRIDLPPVAGFEKDVGRKTNMTVLNMDTPKRLAISRLKNRSLDMYEHRLR
GIAGTVLLANRESQPYLKAFQRYHVSFLLTSASSFTSAEKINPIASKIARKRYMKK
HPSTGMATVLIATTCDRRLFYGFPPFGQDENKRPPIPHYYPDDGIROKPIAKDKMHH
```

CDS

```
/codon_start=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAM18874.1"
/db_xref="GI:20302747"
/translation="MIRGLVCLCLSSAVLAQQGPAYFISVPNLLKVGTEETVSVNIF
NVPNPVRKVYLDYDPDRKSTFSEAEVDVNQDEGLTVVRVNADNLPPESTATQRYVYV
VAKSDDPQLTFQKEAQLVLSFQGGYVFTQDKPIYTPNQVKMRIMPLDQDMTPASQP
VKLEILNPQGIIVEQKTFQGSATGFIATETDFPAPLPGNWTTHAHYGPQMQLNVSTQ
FEVREYVLPITYGVKIIIPSNPYILPQDDVISGEVEALTYGKGVDFDLKLGIIIDQEG
NRQLFAQLQTEVNDGYGYEITQRIKDLDLWFPEGSRLYLEAAVTEAGGLREISVL
TSISFETSPFKIGYDLTATHFKPGLPFLVLTLYTPNPKPAEDIPVRVSATAIIPGQD
PIVILGRNNEHNSDTTNOYQASFTVDVPPATQTLTVAQTEQVGLPAAHQAEDEFA
TPYQSPSASYLLVRVLRGPVDEAIDVEAVTKNDIQSYNMYVTVTRGVTHQCKI
VRQGGVLKTIITFRTSAVMAPISRLIYVYINLQGEVADVSTLEIENVCRNKVTSSVD
NVEPQEQANIEVEADPNLSVGLAVDQAVYLLNNYRLTSQKMFQAMAKYDQCGPGG
GQDSANVFKDAGTVLTNTQLSPAVRGNAGCGTQSRKRDQLQNSQEAQKSLFVCMVIT
VMIVRAKLNSTQDECYFAFLTCNHAARSLRQGRGRMGGGGLLDIPIDEDESQLE
VARTFETPTWIFEDVQDDGGQALVPVTVPGSITTTWVIQAVGISTANGCMVAEPFHE
SFKFFFIHLQLPYSIIRGEQAVRATIFNYDQDLRVNVMQGVGVCSGARAGERSE
RKTLFIKGNDAAASVLFPIIPLFVGTFFIRVLAFTAAAGDIIKSLQVI"
```

mRNA

```
/codon_start=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAM18875.1"
/db_xref="GI:20302748"
/translation="MASETAPVWPQNYTSPDKRKEGGIRHRRLLVFLTLLCLSVITNA
LLWLSRQGLTNMLLARRDKNTDGNLSRPVPEVSEAPFSVEDVELRIRNAASKL
HPDRKFRLGKWPVCLRTGKTKSKKHCPPSKHIGNYRCAVGVNGSGLLGSDCGAD
IDAKDYVIRIDLPPVAGFEKDVGRKTNMTVLNMDTPKRLAISRLKNRSLDMYEHRLR
GIAGTVLLANRESQPYLKAFQRYHVSFLLTSASSFTSAEKINPIASKIARKRYMKK
HPSTGMATVLIATTCDRRLFYGFPPFGQDENKRPPIPHYYPDDGIROKPIAKDKMHH
```

CDS

```
/codon_start=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAM18876.1"
/db_xref="GI:20302749"
/translation="MASETAPVWPQNYTSPDKRKEGGIRHRRLLVFLTLLCLSVITNA
LLWLSRQGLTNMLLARRDKNTDGNLSRPVPEVSEAPFSVEDVELRIRNAASKL
HPDRKFRLGKWPVCLRTGKTKSKKHCPPSKHIGNYRCAVGVNGSGLLGSDCGAD
IDAKDYVIRIDLPPVAGFEKDVGRKTNMTVLNMDTPKRLAISRLKNRSLDMYEHRLR
GIAGTVLLANRESQPYLKAFQRYHVSFLLTSASSFTSAEKINPIASKIARKRYMKK
HPSTGMATVLIATTCDRRLFYGFPPFGQDENKRPPIPHYYPDDGIROKPIAKDKMHH
```

MDTEYSFYKRLNQSSKGVKLHIRECDNQ"

mRNA

```
join(<20714. .20772,23503. .23669,24051. .24251,24572. .24624,
25001. .25018,25219. .25350,25808. .25850,26307. .26394,
26832. .26931,27374. .27604,27949. .28146,28746. .28940,
29352. .29543,30020. .30166,30774. .30848,31042. .31200,
32189. .32403,32972. .33200,33679. .>33891)
/product="unknown"
```

CDS

```
join(20714. .20772,23503. .23669,24051. .24251,24572. .24624,
25001. .25018,25219. .25350,25808. .25850,26307. .26394,
26832. .26931,27374. .27604,27949. .28146,28746. .28940,
29352. .29543,30020. .30166,30774. .30848,31042. .31200,
32189. .32403,32972. .33200,33679. .>33891)
/note="CDS5"
```

/codon\_start=1

/evidence=not\_experimental

/product="unknown"

/protein\_id="AAM18874.1"

/db\_xref="GI:20302747"

```
/translation="MIRGLVCLCLSSAVLAQQGPAYFISVPNLLKVGTEETVSVNIF
NVPNPVRKVYLDYDPDRKSTFSEAEVDVNQDEGLTVVRVNADNLPPESTATQRYVYV
VAKSDDPQLTFQKEAQLVLSFQGGYVFTQDKPIYTPNQVKMRIMPLDQDMTPASQP
VKLEILNPQGIIVEQKTFQGSATGFIATETDFPAPLPGNWTTHAHYGPQMQLNVSTQ
FEVREYVLPITYGVKIIIPSNPYILPQDDVISGEVEALTYGKGVDFDLKLGIIIDQEG
NRQLFAQLQTEVNDGYGYEITQRIKDLDLWFPEGSRLYLEAAVTEAGGLREISVL
TSISFETSPFKIGYDLTATHFKPGLPFLVLTLYTPNPKPAEDIPVRVSATAIIPGQD
PIVILGRNNEHNSDTTNOYQASFTVDVPPATQTLTVAQTEQVGLPAAHQAEDEFA
TPYQSPSASYLLVRVLRGPVDEAIDVEAVTKNDIQSYNMYVTVTRGVTHQCKI
VRQGGVLKTIITFRTSAVMAPISRLIYVYINLQGEVADVSTLEIENVCRNKVTSSVD
NVEPQEQANIEVEADPNLSVGLAVDQAVYLLNNYRLTSQKMFQAMAKYDQCGPGG
GQDSANVFKDAGTVLTNTQLSPAVRGNAGCGTQSRKRDQLQNSQEAQKSLFVCMVIT
VMIVRAKLNSTQDECYFAFLTCNHAARSLRQGRGRMGGGGLLDIPIDEDESQLE
VARTFETPTWIFEDVQDDGGQALVPVTVPGSITTTWVIQAVGISTANGCMVAEPFHE
SFKFFFIHLQLPYSIIRGEQAVRATIFNYDQDLRVNVMQGVGVCSGARAGERSE
RKTLFIKGNDAAASVLFPIIPLFVGTFFIRVLAFTAAAGDIIKSLQVI"
```

BASE COUNT 10151 a 7072 c 7010 g 10311 t  
ORIGIN

Alignment Scores:	592	Length:	34544
Pred. No.:	7.00	Matches:	7
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	43.75%	Indels:	0
Query Match:	3	Gaps:	0
DB:			

US-09-854-133-587 (1-16) x AF391289 (1-34544)

QY 6 GlyIleAspPheIleIlePhe 12

|||||

Db 30200 GGGATAGATTTCATTATTTT 30220

RESULT 43

L16621/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

L16621  
Caenorhabditis elegans cosmid ZK688, complete sequence.  
L16621  
HTG.  
Caenorhabditis elegans.  
Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 37977)  
Waterston,R.  
Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
Science 282 (5396), 2012-2018 (1998)  
99069613  
9851916  
2 (bases 1 to 37977)  
Wilson,R.  
The sequence of C. elegans cosmid ZK688  
Unpublished (2001)

37977 bp DNA linear INV 29-MAY-2002



REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 37977)  
Waterston,R.  
Direct Submission  
Submitted (09-AUG-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
4 (bases 1 to 37977)  
Waterston,R.  
Direct Submission  
Submitted (07-SEP-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
5 (bases 1 to 37977)  
Waterston,R.  
Direct Submission  
Submitted (05-OCT-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
6 (bases 1 to 37977)  
Waterston,R.  
Direct Submission  
Submitted (29-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Sep 7, 2001 this sequence version replaced gi:289775.  
Submitted by:

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all  
regions were double stranded, sequenced with an alternate chemistry  
or covered by high quality data (i.e., phred quality >= 30); an  
attempt was made to resolve all sequencing problems, such as  
compressions and repeats; all regions were covered by sequence from  
more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its  
analysis see:  
<http://www.wormbase.org/db/seq/sequence?name=ZK688;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C02C2, 1600 bp overlap; the 3' cosmid is C29E4,  
1000 bp overlap. Actual start of this cosmid is at base position 1  
of ZK688; actual end is at 36977 of ZK688.

NOTES:

Coding sequences below are the result of integration and manual  
review of the following data : computer analysis using the program  
Genefinder (P. Green and L. Hillier, personal communication), the  
large scale EST projects of Yuji Kohara  
([http://www.ddbj.nig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html)) and The C.  
elegans ORFeome cloning project (<http://worfdb.dfci.harvard.edu/>),  
similarity to other proteins from BlastX analyses  
(<http://blast.wustl.edu/>), sequence conservation with C. briggsae  
using Jim Kent's WABA alignment program (Genome Research  
10:1115-1125, 2000), individual C. elegans GenBank submissions,  
and personal communications with C. elegans researchers. tRNAs  
are predicted using the program tRNAscan-SE (Lowe, T.M. and  
Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).  
Location/Qualifiers  
1..37977  
/organism="Caenorhabditis elegans"

FEATURES  
source

gene

CDS

gene

CDS

/strain="Bristol N2"  
/db\_xref="taxon:6239"  
/chromosome="III"  
/clone="ZK688"  
complement(98..3813)  
/gene="ZK688.9"  
/note="for a graphical representation of this gene see:  
<http://www.wormbase.org/db/seq/sequence?name=ZK688.9;class=Sequence>"  
complement(join(98..296,379..461,1044..1160,1294..1575,  
1684..1860,2714..2837,2886..3052,3097..3201,3778..3813)  
/gene="ZK688.9"  
/note="coded for by the following C. elegans cDNAs:  
yk657d4.3, yk657d4.5"  
/codon\_start=1  
/product="Hypothetical protein ZK688.9"  
/protein\_id="AAL00884.2"  
/db\_xref="GI:15967147"  
/translation="MSAKNRLLDVFDKTRDVIESDLVKMRGYKRNIGSRNQEMTMKNV  
EKKVEQMGAKNEEMFSISKIMQKPNSTGHLQWDSQTLADLMNKSDSFDQDQESLNMP  
RSSESSAMSVRNRLERQSEPVKSPKAAKPKVKNAGITASYVTSEMSDRRSIDEVMV  
RAAKNAKREKFDLGFKFISLAGHILESSCKHSEDOEDPSCLKCKYDRELKLEERPE  
MVFARNSLTIQFGLGSIIEFNALDALKMVCADRLPDVKVGASTVWQSARQDRIOQISE  
HOKPFDWTYTHYKGTGCVQVTPPTTERIDMERLKRDEILFSSSITLFEDELADHGI  
AQLARVRMRGYFFVLLRFYMRVDNVLRLVCDTRIVGNEFDGHVIREWQLREAKYGN  
LGHVDPEELLVDRAWMHLPVVEEHFDRVSVDRERLF"  
complement(4277..10419)  
/gene="ZK688.5"  
/note="for a graphical representation of this gene see:  
<http://www.wormbase.org/db/seq/sequence?name=ZK688.5;class=Sequence>"  
complement(join(4277..4461,4657..4827,4889..5122,  
5283..5975,6021..6213,6638..8566,8618..9006,9053..9315,  
9416..9622,9674..9985,10085..10251,10300..10419)  
/gene="ZK688.5"  
/note="contains similarity to Pfam domain PF00240  
(ubiquitin), Score=14.4, E-value=0.012, N=1; coded for by  
the following C. elegans cDNAs: yk16d10.3, yk16d10.5,  
yk36d7.3, yk36d7.5, yk39d2.3, yk39d2.5, yk153b5.5,  
yk16la3.5, yk171c4.5, yk171c4.3, yk493c6.5, yk784g03.3"  
/codon\_start=1  
/product="Hypothetical protein ZK688.5"  
/protein\_id="AAL00883.1"  
/db\_xref="GI:15487652"

/translation="MSETQEASNGEPDLPTTIRVTLKLLDDREATVTIGLQDTIQ  
SLIDLGRREMNIQSGFQVIAGGRVLNSTQTVAAGISDGTQVHLVDKSGSGENDRPN  
VMPDRVAGPRIINAIPGLPPGFIFQSPAFARMIPGNVEIPTPPSQTHVHPVRVP  
GSIAGTPVLTSRVSEDCVLOKAVPYRQTSNPNRPAASQTVTFPPSNVIOQWTVNIA  
DDLIFRPREHFEQVRETINNISFLSDSTRLGVMKNQNSQTSLSVELPPVSPHISP  
ALEKLDLCLMTDHLSTRFDKLEHDLVAATHRVLEVMVKTRQFDQNLQQARQVE  
ALDEIVKHLEYQWAELESHMKDFERIRFNKQTEERALKIQEYPETPRNPHYMRHAL  
DIDVIGVMREFKQQRTRLEDLLDLNEVGAVKFIIRDHITTEVDYRYQALSMFYCY  
IQMRHQISHMTHLTADLDVSFITSPFPQRIPLQYAFQSSDEQPLTIPVTHIFPPRM  
LSDVGGRYNGDYPYLAHPSPVHMEVVLQEPRIAEPRPQILARPTPQQFVLTQEMNQ  
GGSINLMAATDPIISGLSLQDVLRAQQEQVENLFAQQPGIEGGRVVRTARPPRRVATT  
GDAPSPVETLPPGSDQQPGTSRRFTTHRFNVQPDARGAETETLAPFPVAIEPNELQ  
RIAKNIASRYRAEALQRIASSLTTRFNESWDTRIQNMPCLTLRECVAIAMLEMTSAG  
NTVNESKDLMLALVRDEEVLVRAIECKISLFGREFPTHVARLIMPTNQATASRNDY  
ESVDGVEESQSDDEFDSMIVRPSDISQPSQGLRAIRERRQNRQFLENRGRIPTSS  
APSTSENPPGPFNSEDAADIRAGRLPIGTRPNRRTVRETIVHPAAARAEASNHISLT  
FTATHTFAPAGFPLMMASSNVPSISAGPPGWIQVSPVPTTRTGLFDFLSSGSDQ  
PARSTPPAPTPTTSAPATVQSSPTTRQESMDIDSPNVQPNPGHVESPAIAAQAAR  
VARARIDHLAATFNGDLADSRQSPFVTPGPTPLNDPRRTVVRTVQHKPMVAIDPF  
MNCTNRHCEINRLATPTHMADGDRFTLQSLQPDQEEFARIQALVFSIERRPIQHED  
EDYNSIRRTQSGLLSFRNLEDFRPVKTAIRSLLAHCIDADTLFTMNMNNSIGYQAA  
NHTELLRMVKEHISRVPGSRATRNASQNTTSVNSQSTETSPREQMNRSPVQEAADPRLA  
FSPFPFGINLEQEVLIQGLIASLLTYLDVYMESSNPRPPGIFGLLELTGRLTRHD  
FAQLARRTTATNVASEFEACIRAHIRDNYLVGRTGLSNSELHGAENLANNEQFFAIF  
MSQNDQLPTSPFPFGYRNDFAEVVWAFROIETALIKSLFSLQNLDSGNVRFILQS  
VDSYLIRNLMIFMYCRDVRMVKIQMRISDYFATIRSEYLSLQNLDSGNVRFILQS  
DYWSNRYTDFSENFQFLLKVRAGTDWNDIVLNSQLLPTIASTSSQPTSSSSSLN  
TVSTNNPRKRNHGSDREDNDDCVDVAVPMMTSLSTSSPLTSQSSSGSTRTSSGSSGSP  
TSSTTTNNIQ"



gene 12630..13366  
/gene="ZK688.3"  
/note="for a graphical representation of this gene see:  
http://www.wormbase.org/db/seq/sequence?name=ZK688.3;class  
-Sequence"  
join(12630..12752,12797..12904,12953..13366)  
/gene="ZK688.3"  
/note="contains similarity to Pfam family PF01557  
(Fumarylacetoacetate (FAA) hydrolase family), score=340.3,  
E=2.1e-98, N=1; coded for by the following C. elegans  
cDNAs: yk899e08.5, CEESF53F, CEESP96F, yk899e08.3"  
/codon\_start=1  
/product="Hypothetical protein ZK688.3"  
/protein\_id="AAA28230.1"  
/db\_xref="GI:289782"  
/translation="MSSLAGFRNLATKIVCVGRNYKDHAELELGNAPKKPMLFVKTVN  
SPVEGEPIVAPPGCNLHQEVELGVVISKASRISKSDAMDYIGGYTVALDMTARDF  
QDEAKKAGAPWFLAKSFDGSCPIGGFLPVS DIPNPHDVELFCKINGKQQRCRTDVM I  
FDIPTLLEYTTQFTTLEVGVDVLTGTGTPAGVTKINSGDVIEFGLTDKLSKFNVO"  
13883..14129  
/gene="ZK688.4"  
/note="for a graphical representation of this gene see:  
http://www.wormbase.org/db/seq/sequence?name=ZK688.4;class  
-Sequence"  
join(13883..14027,14071..14129)  
/gene="ZK688.4"  
/codon\_start=1  
/product="Hypothetical protein ZK688.4"  
/protein\_id="AAA28229.1"  
/db\_xref="GI:289781"  
/translation="MHTLILFRVHVSNSFYIFSSTILLTSAVILAIFFFFGDGFRE  
KYRREFSTNSKEKDGKNLTTVE"  
19891..19892  
/note="SL1 trans-splice acceptor; see yk729b7.5"

misc\_feature

Alignment Scores:  
Pred. No.: 644 Length: 37977  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 3 Gaps: 0

US-09-854-133-587 (1-16) x L16621 (1-37977)  
QY 6 GlyIleAspPheIleIlePhe 12  
|||||  
Db 17672 GGGATTGACTTATCATTTT 17652

RESULT 44  
AC002499 41052 bp DNA linear PRI 10-OCT-1997  
LOCUS Human Cosmid g5129s435 from 7q31.3, complete sequence.  
DEFINITION AC002499  
ACCESSION AC002499  
VERSION AC002499.1 GI:2342698  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 41052)  
AUTHORS Iadonato,S.P., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D.,  
Green,P. and Olson,M.V.  
TITLE Large-scale MCD Mapping and Sequencing of Human Chromosome 7  
JOURNAL Unpublished (1996)  
REMARK University of Washington Human Genome Center  
Box 352145 Seattle, WA 98195  
Contact: Shawn Iadonato (iadonato@u.washington.edu)  
2 (bases 1 to 41052)  
AUTHORS Iadonato,S.P., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D.,  
Green,P. and Olson,M.V.  
TITLE Direct Submission  
JOURNAL Submitted (25-AUG-1997) Human Genome Center, University of

REMARK Washington, Box 352145, Seattle, WA 98195, USA  
University of Washington Human Genome Center  
Box 352145 Seattle, WA 98195  
Contact: Shawn Iadonato (iadonato@u.washington.edu)

REFERENCE 3 (bases 1 to 41052)  
AUTHORS Iadonato,S.P., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D.,  
Green,P. and Olson,M.V.  
TITLE Direct Submission  
JOURNAL Submitted (10-OCT-1997) Human Genome Center, University of  
Washington, Box 352145, Seattle, WA 98195, USA  
REMARK University of Washington Human Genome Center  
Box 352145 Seattle, WA 98195  
Contact: Shawn Iadonato (iadonato@u.washington.edu)

COMMENT Overlapping Sequences:  
5': UWGC:g51292l01  
3': UWGC:g5129g124

Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.

Pct of Consensus above quality 40: 97.9%  
Number of manually edited bases: 7

Double stranded (DS) coverage: 91.0%  
DS or two chemistry coverage: 7.0%  
Single stranded regions: 3

Quality Cumulative Percentage of Consensus  
-----  
90 xx ( 8.3%)  
80+ xxxxxxxx (40.6%)  
70+ xxxxxxxxxxxxxxxxxx (76.2%)  
60+ xxxxxxxxxxxxxxxxxxxxxx (88.3%)  
50+ xxxxxxxxxxxxxxxxxxxxxx (94.3%)  
40+ xxxxxxxxxxxxxxxxxxxxxx (97.9%)  
30+ xxxxxxxxxxxxxxxxxxxxxx (99.6%)  
20+ xxxxxxxxxxxxxxxxxxxxxx (100.0%)  
10+ xxxxxxxxxxxxxxxxxxxxxx (100.0%)  
00+ xxxxxxxxxxxxxxxxxxxxxx (100.0%)

Base-by-base quality values are not generally visible from the  
Genbank flat file format but are available as part  
of this entry's ASN.1 file.

Sequence Validation:  
This sequence has been validated by Multiple Complete Digest  
Mapping. Comparison of the experimentally derived map digest  
fragments with sequence-predicted fragments is given below.  
Small fragments below a variable cutoff (approximately 400-600bp)  
are not mapped and hence do not appear in the table. There are no  
significant remaining discrepancies between the experimental and  
predicted values. Uniquely ordered fragment groups are separated  
by dashed lines.

ECORI		HindIII		NsiI	
Map	Seq	Map	Seq	Map	Seq
3714.42	3709.00	1212.81	1206.00	11295.40	11398.00
5425.43	5420.00	1591.62	1586.00	2802.45	2811.00
13174.96	13413.00	1902.56	1912.00	16219.29	16613.00
8209.13	8223.00	10264.00	10325.00	3557.24	3563.00
		3745.48	3711.00	5799.89	5846.00
		4214.65	4237.00		
		1794.23	1814.00		

```

/rpt_family="L1"
complement(37523. .39574)
/rpt_family="L1"

BASE COUNT    13425 a    7256 c    7164 g    13207 t
ORIGIN

Alignment Scores:
Pred. No.:      691
Score:          7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match:    43.75%
DB:             9

US-09-854-133-587 (1-16) x AC002499 (1-41052)

QY    9  PheillePheTrpIlePhe 15
      |||||
Db    40866 TTTATCATATTTGGATATT 40886

RESULT 45
AC003032
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
Location/Qualifiers
1. .42354
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="Y"
/cell_line="human/hamster J640-51"
/clone_lib="LLOYNCO3 'M'"
/note="The library was constructed by J.
Jong at Lawrence Livermore National Labora
flow-sorted Y chromosomes. Vector: Lauri
BASE COUNT    13578 a    8319 c    8145 g    12312 t
ORIGIN

Alignment Scores:
Pred. No.:      710
Score:          7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match:    43.75%
DB:             9

US-09-854-133-587 (1-16) x AC003032 (1-42354)

Qy    9  PheillePheTrpIlePhe 15

```

us-09-854-133-587.oligo.rge

• Tue May 13 12:12:57 2003

Db 31661 TTTATTTCTCTGGATTTT 31681

Search completed: May 11, 2003, 17:20:51  
Job time : 424.071 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 16:14:30 ; Search time 34.2655 Seconds  
(without alignments)  
1051.553 Million cell updates/sec

Title: US-09-854-133-587

Perfect score: 16

Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4348058

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 60 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_174132\_694/app\_query.fasta\_1.462  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdl  
-LIST=60 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=45 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn\_1\_1\_275\_@runat\_05052003\_174132\_694 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	16	100.0	337	23	AAD23462	Human lung tumour-
2	16	100.0	2239	23	AAD23460	Human lung tumour-
3	16	100.0	5981	23	AAD23461	Human lung tumour-
C	7	43.8	579	24	ABN65138	Human cancer relat
C	5	7	826	16	AAQ79736	Flower style-speci
C	6	7	826	18	AAT64553	Tomato S-ribonucle
C	7	7	2241	18	AAT64552	Tomato S-ribonucle
8	7	43.8	5886	24	ABL34212	Human immune syste
9	7	43.8	8576	24	ABL34229	Human immune syste
10	7	43.8	9265	24	ABL34153	Human immune syste
11	7	43.8	9265	24	AAS63355	Chemically pretrea
12	7	43.8	15723	23	ABL19746	Drosophila melanog
C	13	7	15881	24	ABL70143	Chemically treated
14	7	43.8	15881	24	ABL32260	Human immune syste
15	7	43.8	15881	24	AAS61069	Human gene regulat
16	7	43.8	15881	24	ABK31182	Signal transductio
17	7	43.8	16633	24	ABN79984	Human chemically m
18	6	37.5	26	19	AAV42717	3' PCR primer used
C	19	6	246	21	AAC30925	Human secreted pro
C	20	6	250	24	ABL77073	Human ovarian canc
21	6	37.5	277	22	AAS47046	Human breast cance
22	6	37.5	277	22	AAF17616	Human breast cance
23	6	37.5	277	24	ABK95081	Human breast tumou
24	6	37.5	283	24	ABL87080	Human ovarian canc
C	25	6	303	22	AAK77254	Human immune/haema
C	26	6	312	23	ABV17400	Human prostate exp
C	27	6	339	23	ABV47195	Human prostate exp
C	28	6	416	23	AAS66468	DNA encoding novel
29	6	37.5	418	20	AAV90289	EST clone DK70. H
30	6	37.5	498	24	ABN94582	Gene #1080 used to
31	6	37.5	530	24	ABQ35704	Oligonucleotide fo
C	32	6	530	24	ABQ35705	Oligonucleotide fo
33	6	37.5	548	22	AAH11384	Human cDNA clone (
34	6	37.5	553	24	ABQ45838	Oligonucleotide fo
C	35	6	553	24	ABQ45839	Oligonucleotide fo
36	6	37.5	559	24	ABQ44134	Oligonucleotide fo
C	37	6	559	24	ABQ44135	Oligonucleotide fo
C	38	6	572	22	AAD05092	Human secreted pro
39	6	37.5	579	22	AAAL34901	Human musculoskele
40	6	37.5	582	23	AAS85327	DNA encoding novel
C	41	6	584	21	AAC08897	Human secreted pro
42	6	37.5	600	22	ABA62551	Human foetal liver
43	6	37.5	600	22	ABA29879	Probe #8345 for ge
44	6	37.5	600	22	AAK10907	Human brain expres
45	6	37.5	600	22	AAK36765	Human bone marrow
46	6	37.5	600	22	AAI17617	Probe #7550 for ge
47	6	37.5	600	22	AAI42536	Probe #11222 used
48	6	37.5	600	24	ABS10768	Human genome-deriv
49	6	37.5	675	24	ABQ56800	Human colon cancer
50	6	37.5	696	22	ABA01232	GIP6-22 coding seq
C	51	6	696	24	AAH73248	Bacillus clausii g
C	52	6	740	22	AAH73248	Human cervical can
53	6	37.5	748	22	AAI95673	Human neuroblastom
54	6	37.5	764	24	ABQ24630	Oligonucleotide fo
C	55	6	764	24	ABQ24631	Oligonucleotide fo
C	56	6	801	22	AAH00733	Clostridium tetani
C	57	6	802	22	AAH04913	Human cDNA clone (
C	58	6	822	21	AAA70240	Plasmodium falcipa
59	6	37.5	831	24	ABQ15112	Oligonucleotide fo
C	60	6	831	24	ABQ15113	Oligonucleotide fo

ALIGNMENTS

RESULT 1  
AAD23462  
ID AAD23462 standard; cDNA; 337 BP.

```
XX AAD23462;
AC 26-FEB-2002 (first entry)
DT
XX Human lung tumour-specific 20E10 5' CDNA.
DE
XX Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW antisense-therapy; vaccine; immune response; lung cancer; 20E10; ss.
XX Homo sapiens.
OS
XX WO200172295-A2.
PN
XX 04-OCT-2001.
PD
XX 28-MAR-2001; 2001WO-US09991.
PF
XX 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX (CORI-) CORIXA CORP.
PA
XX Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX WPI; 2001-639201/73.
DR
XX New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer -
PT Claim 1; Page 334; 378pp; English.
XX The invention relates to isolated lung tumour-specific proteins and
CC their corresponding CDNA molecules. Lung tumour-specific proteins and
CC their antigen-presenting cells are useful for stimulating and/or
CC expanding T cells specific for a tumour protein, and for inhibiting
CC the development of cancer. The invention also relates to a composition
CC useful for stimulating an immune response, and for treating cancer. The
CC lung tumour specific oligonucleotide is useful in gene therapy and for
CC diagnosis, detection and treatment of lung cancer. The present sequence
CC is a CDNA encoding human lung tumour-specific protein.
XX Sequence 337 BP; 103 A; 60 C; 93 G; 81 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 1.79e-09 Length: 337
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0
US-09-854-133-587 (1-16) x AAD23462 (1-337)
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
Db 107 TTCCAGGCCAATTGTGGCATAGATTTTATCATATCTCGATTTTGG 154
RESULT 2
AAD23460
ID AAD23460 standard; CDNA; 2239 BP.
XX
AC AAD23460;
XX 26-FEB-2002 (first entry)
DT
XX Human lung tumour-specific 19A4 CDNA.
DE
XX
```

```
KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW antisense-therapy; vaccine; immune response; lung cancer; 19A4; ss.
XX Homo sapiens.
OS
XX WO200172295-A2.
PN
XX 04-OCT-2001.
PD
XX 28-MAR-2001; 2001WO-US09991.
PF
XX 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX (CORI-) CORIXA CORP.
PA
XX Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX WPI; 2001-639201/73.
DR
XX New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer -
PT Claim 1; Page 332; 378pp; English.
XX The invention relates to isolated lung tumour-specific proteins and
CC their corresponding CDNA molecules. Lung tumour-specific proteins and
CC their antigen-presenting cells are useful for stimulating and/or
CC expanding T cells specific for a tumour protein, and for inhibiting
CC the development of cancer. The invention also relates to a composition
CC useful for stimulating an immune response, and for treating cancer. The
CC lung tumour specific oligonucleotide is useful in gene therapy and for
CC diagnosis, detection and treatment of lung cancer. The present sequence
CC is a CDNA encoding human lung tumour-specific protein.
XX Sequence 2239 BP; 619 A; 444 C; 493 G; 683 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 9.06e-09 Length: 2239
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0
US-09-854-133-587 (1-16) x AAD23460 (1-2239)
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
Db 104 TTCCAGGCCAATTGTGGCATAGATTTTATCATATCTCGATTTTGG 151
RESULT 3
AAD23461
ID AAD23461 standard; CDNA; 5981 BP.
XX
AC AAD23461;
XX 26-FEB-2002 (first entry)
DT
XX Human lung tumour-specific 14F10 full length CDNA.
DE
XX Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW antisense-therapy; vaccine; immune response; lung cancer; 14F10; ss.
XX Homo sapiens.
OS
XX WO200172295-A2.
PN
XX
```

PD 04-OCT-2001.  
XX  
PF 28-MAR-2001; 2001WO-US09991.  
XX  
XX 29-MAR-2000; 2000US-0538037.  
PR 05-JUN-2000; 2000US-0588937.  
PR 18-AUG-2000; 2000US-0640878.  
PR 22-SEP-2000; 2000US-234517P.  
PR 01-NOV-2000; 2000US-0704512.  
PR 14-DEC-2000; 2000US-0738973.  
XX

XX (CORI-) CORIXA CORP.  
PA

PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;  
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
XX WPI; 2001-639201/73.  
DR

XX New human lung-specific polynucleotides and polypeptides for the  
PT diagnosis and treatment of disease e.g. lung cancer -  
PT

PS Claim 1; Page 332-334; 378pp; English.  
XX

CC The invention relates to isolated lung tumour-specific proteins and  
CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
CC their antigen-presenting cells are useful for stimulating and/or  
CC expanding T cells specific for a tumour protein, and for inhibiting  
CC the development of cancer. The invention also relates to a composition  
CC useful for stimulating an immune response, and for treating cancer. The  
CC lung tumour specific oligonucleotide is useful in gene therapy and for  
CC diagnosis, detection and treatment of lung cancer. The present sequence  
CC is a cDNA encoding human lung tumour-specific protein.  
XX

XX SQ Sequence 5981 BP; 1858 A; 1029 C; 1098 G; 1996 T; 0 other;

Alignment Scores:  
Pred. No.: 2.1e-08 Length: 5981  
Score: 16.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x AAD23461 (1-5981)

Qy 1 PheGlnAlaAsnCysGlyIleAaspPheIleIlePheTrpIlePheTrp 16  
Db 102 TTCCAGGCCAATTGTGGCATAGATTTTATCATATCTCGATTTTGG 149

RESULT 4

ABN65138/c

ID ABN65138 standard; cDNA; 579 BP.

XX AC ABN65138;

XX 28-JUN-2002 (first entry)

DE Human cancer related polynucleotide SEQ ID NO 5105.

XX Human; cytostatic; gene expression; gene mapping; tissue profiling;  
KW gene therapy; cancer; tumour; gene; ss.  
XX

OS Homo sapiens.

XX WO200214500-A2.

PN 21-FEB-2002.

XX 16-AUG-2001; 2001WO-US25840.

PF 16-AUG-2000; 2000US-226326P.

XX (CHIR) CHIRON CORP.  
PA

PA (HYSE-) HYSEQ INC.

XX Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;  
PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;  
XX WPI; 2002-241905/29.

DR New nucleic acid for producing a polypeptide, detecting differentially

XX expressed genes correlated with a cancerous state of a mammalian cell,  
PT and inhibiting tumor growth -  
PT

XX Claim 1; SEQ ID NO 5105; 883pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide (ABN27253-ABN33262)  
CC with cytostatic activity. The polynucleotide is used to produce a  
CC polypeptide, to detect differentially expressed genes correlated with a  
CC cancerous state of a mammalian cell and to inhibit tumour growth. The  
CC polynucleotide is used as a probe in mapping and tissue profiling. The  
CC encoded polypeptide and antibodies to the polypeptide can also be used  
CC for therapeutic and diagnostic purposes. The polynucleotide is useful for  
CC gene therapy.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 579 BP; 195 A; 123 C; 100 G; 161 T; 0 other;

Alignment Scores:  
Pred. No.: 22.5 Length: 579  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABN65138 (1-579)

Qy 10 IleIlePheTrpIlePheTrp 16

Db 328 ATAATATTTTGGATATTTGG 308

RESULT 5

AAQ79736/c

ID AAQ79736 standard; cDNA; 826 BP.

XX AC AAQ79736;

XX 21-SEP-1995 (first entry)

DE Flower style-specific tomato S-ribonuclease.

XX S-ribonuclease; flower style-specific; self-incompatibility; ds.

OS Lycopersicon peruvianum.

XX Key Location/Qualifiers

FT CDS 1..489

FT /\*tag= a

FT /note= "START codon absent"

XX JP06335389-A.

XX 06-DEC-1994.

XX 27-MAY-1993; 93JP-0126286.

XX 27-MAY-1993; 93JP-0126286.

XX (KAGO) KAGOME KK.

XX WPI; 1995-056901/08.

XX P-PSDB; AAR66604.

PT Flower style-specific S-ribonuclease and DNA coding sequence -  
PT isolated from self-incompatible wild tomato species  
XX  
PS Claim 4; Pages 5-6; 10pp; Japanese.  
XX  
CC AAQ79736 encodes AAR66604 a flower style-specific S-ribonuclease,  
CC isolated from a self-incompatible wild tomato species.  
XX  
SQ Sequence 826 BP; 309 A; 150 C; 131 G; 236 T; 0 other;

Alignment Scores:  
Pred. No.: 30.5 Length: 826  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 43.75%  
DB: 16 Indels: 0  
Gaps: 0

US-09-854-133-587 (1-16) x AAQ79736 (1-826)

QY 6 GlyIleAspPheIleIlePhe 12  
Db 550 GGAATTGATTTTATATATT 530

RESULT 6  
AAT64553/c  
ID AAT64553 standard; CDNA to mRNA; 826 BP.  
XX  
AC AAT64553;

DT 22-MAY-1997 (first entry)  
XX  
DE Tomato S-ribonuclease cDNA fragment.  
XX  
KW S-ribonuclease; plant style; promoter; tissue-specific;  
KW self-incompatibility; ds.  
XX  
OS Lycopersicon peruvianum.

XX  
FH Key Location/Qualifiers  
FT CDS 1..489  
FT /\*tag= a

XX  
PN JP09028381-A.  
XX  
PD 04-FEB-1997.  
XX  
PF 24-JUL-1995; 95JP-0187557.  
XX  
PR 24-JUL-1995; 95JP-0187557.

XX  
PA (CHEN/) CHENG E K.  
PA (MINA/) MINAMI H.  
PA (TAKA/) TAKAGI M.

XX  
DR WPI; 1997-159091/15.  
DR P-PSDB; AAW14912.

XX  
PT Tomato S-ribonuclease gene and promoter sequence - useful for  
PT expressing genes in plant style cells, e.g. to confer  
PT self-incompatibility on the plant

XX  
PS Claim 4; Page 10-11; 15pp; Japanese.

XX  
CC The present sequence encodes part of a tomato (Lycopersicon  
CC peruvianum) S-ribonuclease. The S-gene can confer self-  
CC incompatibility on a plant.

XX  
SQ Sequence 826 BP; 309 A; 150 C; 131 G; 236 T; 0 other;

Alignment Scores:  
Pred. No.: 30.5 Length: 826  
Score: 7.00 Matches: 7

Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 43.75%  
DB: 18 Indels: 0  
Gaps: 0

US-09-854-133-587 (1-16) x AAT64553 (1-826)

QY 6 GlyIleAspPheIleIlePhe 12  
Db 550 GGAATTGATTTTATATATT 530

RESULT 7  
AAT64552/c  
ID AAT64552 standard; DNA; 2241 BP.  
XX  
AC AAT64552;

XX  
DT 22-MAY-1997 (first entry)  
XX  
DE Tomato S-ribonuclease gene S12 allele.

XX  
KW S-ribonuclease; plant style; promoter; tissue-specific;  
KW self-incompatibility; ds.  
XX  
OS Lycopersicon peruvianum.

XX  
FH Key Location/Qualifiers  
FT promoter 1..1365  
FT /\*tag= a

FT TATA\_signal  
FT /\*note= "plant style-specific"  
FT 1255..1258  
FT /\*tag= b

FT CDS 1366..1596  
FT /\*tag= c

FT intron 1597..1687  
FT /\*tag= d

FT CDS 1688..2101  
FT /\*tag= e

XX  
PN JP09028381-A.  
XX  
PD 04-FEB-1997.

XX  
PF 24-JUL-1995; 95JP-0187557.

XX  
PR 24-JUL-1995; 95JP-0187557.

XX  
PA (CHEN/) CHENG E K.

PA (MINA/) MINAMI H.

PA (TAKA/) TAKAGI M.

XX  
DR WPI; 1997-159091/15.

DR P-PSDB; AAW14911.

XX  
PT Tomato S-ribonuclease gene and promoter sequence - useful for  
PT expressing genes in plant style cells, e.g. to confer  
PT self-incompatibility on the plant

XX  
PS Claim 2; Page 9-10; 15pp; Japanese.

XX  
CC The present sequence is the tomato S12 gene from Lycopersicon  
CC peruvianum coding for an S-ribonuclease; the S12 promoter region  
CC has activity in plant style tissue. The S-gene can confer self-  
CC incompatibility on a plant. In addition, a desired DNA sequence  
CC can be expressed specifically in style cells of a plant using the  
CC S12 promoter sequence.

XX  
SQ Sequence 2241 BP; 683 A; 390 C; 444 G; 724 T; 0 other;

Alignment Scores:  
Pred. No.: 71.7 Length: 2241  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00%  
Conservative: 0



Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 18 Gaps: 0

US-09-854-133-587 (1-16) x AATG4552 (1-2241)

QY 6 GlyIleAspPheIleIlePhe 12  
DB 2203 GGAATTGATTTTATATATTT 2183

RESULT 8  
ABL34212  
ID ABL34212 standard; DNA; 5886 BP.  
XX  
AC ABL34212;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human immune system associated gene SEQ ID NO: 2185.  
XX

KW Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX

OS Homo sapiens.  
XX  
PN WO200200928-A2.  
XX  
PD 03-JAN-2002.  
XX

PF 02-JUL-2001; 2001WO-EP07537.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX

PA (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-130909/17.  
XX

PT Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX

PS Claim 1; SEQ ID NO 2185; 32pp + Sequence Listing; German.

XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX

SQ Sequence 5886 BP; 1482 A; 47 C; 1292 G; 3065 T; 0 other;

Alignment Scores:  
Pred. No.: 164 Length: 5886  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABL34212 (1-5886)

QY 6 GlyIleAspPheIleIlePhe 12  
DB 4386 GGGATTGATTTTATAATTTT 4406

RESULT 9  
ABL34229  
ID ABL34229 standard; DNA; 8576 BP.  
XX  
AC ABL34229;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human immune system associated gene SEQ ID NO: 2202.  
XX

KW Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX

PS Claim 1; SEQ ID NO 2202; 32pp + Sequence Listing; German.

XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX

SQ Sequence 8576 BP; 2122 A; 64 C; 1723 G; 4667 T; 0 other;

Alignment Scores:  
Pred. No.: 226 Length: 8576  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABL34229 (1-8576)

QY 6 GlyIleAspPheIleIlePhe 12

DB 5203 GGAATAGATTTTATAATTTT 5223

RESULT 10  
ABL34153

ID XX ABL34153 standard; DNA; 9265 BP.  
AC ABL34153;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human immune system associated gene SEQ ID NO: 2126.  
XX  
KW Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200200928-A2.  
XX  
PD 03-JAN-2002.  
XX  
PF 02-JUL-2001; 2001WO-EP07537.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-130909/17.  
XX  
PT Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
PS Claim 1; SEQ ID NO 2126; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 9265 BP; 2686 A; 133 C; 1801 G; 4645 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 242 Length: 9265  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0  
  
US-09-854-133-587 (1-16) x ABL34153 (1-9265)  
  
QY 9 PheillelPheTrpIlePhe 15  
|||||  
Db 525 TTTATTATTTTTGGATATTT 545  
  
RESULT 11  
AAS63355  
ID AAS63355 standard; DNA; 9265 BP.  
XX  
AC AAS63355;  
XX  
DT 29-JAN-2002 (first entry)  
XX

DE Chemically pretreated metabolism associated gene #50.  
XX  
KW Human; cytostatic; anti-tumour; metabolism; metabolic disease; liver;  
KW solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;  
KW single nucleotide polymorphism detection; SNP; stool; urine; lung;  
KW cerebral-spinal fluid; intestine; brain; heart; prostate; breast;  
KW DUSP2; EPHX2; QDPR; SGSH; SHMT2; SLC7A2; SLC7A4; TYMS; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200176451-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 06-APR-2001; 2001WO-EP04016.  
XX  
PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-010834/01.  
XX  
PT New nucleic acid, useful for diagnosis and therapy of metabolic  
PT disease, solid tumour and cancers, comprises segment of chemically  
PT modified genomic sequences of genes associated with metabolism -  
XX  
PS Claim 1; Page 129-132; 143pp; English.  
XX  
CC The invention relates to a nucleic acid (I) comprising a sequence at  
CC least 18 bases of a segment of the chemically pretreated DNA of genes  
CC associated with metabolism such as DUSP2 (NM\_004418), EPHX2 (NM\_001979),  
CC QDPR (NM\_000320), SGSH (NM\_000199), SHMT2 (NM\_005412), SLC7A2  
CC (NM\_003046), SLC7A4 (NM\_004173) and TYMS (NM\_001071) (all  
CC undefined). (I) are useful for diagnosis and therapy of metabolic  
CC disease, solid tumours and cancers; as primer oligonucleotides for the  
CC amplification of DNA sequences, for detecting the cytosine methylation  
CC state and/or single nucleotide polymorphisms (SNPs) in a chemically  
CC treated DNA of genes associated with metabolism. An array of (I) is  
CC useful for ascertaining genetic and/or epigenetic parameters for the  
CC diagnosis and/or therapy of existing diseases or the predisposition to  
CC specific diseases by analysing cytosine methylations. The method involves  
CC chemically treating genomic DNA sample by a solution of bisulphite,  
CC hydrogen sulphite or disulphite such that cytosine bases which are  
CC unethylated at the 5th-position are converted to uracil or another base  
CC which is dissimilar to cytosine in terms of hybridisation behaviour and  
CC amplifying fragments of the chemically pretreated genomic DNA. The  
CC genomic DNA is from cells or cellular components which contain DNA,  
CC sources of DNA comprising, for e.g. cell lines, biopsies, blood, sputum,  
CC stool, urine, cerebral-spinal fluid, tissue embedded in paraffin such as  
CC tissue from eye, intestine, kidney, brain, heart, prostate, lung, breast  
CC or liver, histologic object slides and their combinations. Genetic  
CC parameters are mutations, in particular insertions, deletions, point  
CC mutations, inversions and polymorphisms of genes associated with  
CC metabolism and sequences further required for their regulation.  
CC Epigenetic parameters are in particular cytosine methylations and  
CC further chemical modifications of DNA bases of genes associated with  
CC metabolism. Further epigenetic parameters include for e.g. the  
CC acetylation of histones which correlates with DNA methylation.  
CC AAS63306-AAS63373 represent chemically pretreated metabolism associated  
CC genes, and related primers of the invention.  
XX  
SQ Sequence 9265 BP; 2686 A; 133 C; 1801 G; 4645 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 242 Length: 9265  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x AAS63355 (1-9265)

QY 9 PheIleIlePheTrpIlePhe 15  
| | | | | | | | | | | | | | |  
Db 525 TTTATTATTTTGGATATTT 545

RESULT 12

ABL19746/C  
ID ABL19746 standard; DNA; 15723 BP.

XX AC ABL19746;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 10711.

XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

PS Claim 1; SEQ ID NO 10711; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 15723 BP; 5347 A; 2844 C; 2677 G; 4855 T; 0 other;

Alignment Scores:

Pred. No.: 380 Length: 15723  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x ABL19746 (1-15723)

QY 9 PheIleIlePheTrpIlePhe 15

| | | | | | | | | | | | | | |

Db 13139 TTCATAATTTTGGATATTT 13119

RESULT 13  
ABL70143  
ID ABL70143 standard; DNA; 15881 BP.

XX ABL70143;

XX 01-JUL-2002 (first entry)

XX Chemically treated cell signalling DNA sequence#17.

XX Cell signalling; cytosine methylation; cell signalling disease;  
KW cancer; tumour; cytostatic; ds.

XX Unidentified.

XX WO200202807-A2.

XX 10-JAN-2002.

XX 29-JUN-2001; 2001WO-EP07471.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-154758/20.

XX Nucleic acid, useful for diagnosis and therapy of diseases associated  
PT with cell signalling e.g. cancer, comprises chemically modified genomic  
PT sequences of genes associated with cell signalling -

XX Claim 1; SEQ ID NO 33; 24pp+sequence listing; English.

XX The invention relates to a nucleic acid comprising a sequence of at least  
CC 18 bases of a segment of chemically pretreated DNA of genes associated  
CC with cell signalling. The activity of the modified sequences of the  
CC invention may be described as cytostatic. The object of the invention is  
CC to provide the chemically modified DNA of genes associated with cell  
CC signalling, as well as oligonucleotides and/or PNA-oligomers for  
CC detecting cytosine methylations, as well as a method which is  
CC particularly suitable for the diagnosis and/or therapy of genetic and  
CC epigenetic parameters of genes associated with cell signalling. The  
CC chemically modified DNA provided by the invention is useful for diagnosis  
CC and therapy of diseases such as solid tumours and cancer. The sequences  
CC given in records ABL70111-ABL70626 represent chemically pre-treated  
CC genomic DNA's of genes associated with cell signalling.

CC Note: The sequence data for this patent is not represented in the printed  
CC specification, but is based on sequence information supplied by the  
CC European Patent Office.

XX Sequence 15881 BP; 4515 A; 204 C; 3588 G; 7574 T; 0 other;

Alignment Scores:

Pred. No.: 384 Length: 15881  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABL70143 (1-15881)

QY 10 IleIlePheTrpIlePheTrp 16

| | | | | | | | | | | | | | |

Db 11862 ATTATTTTGGATATTTGG 11882

RESULT 14

ABL32260

ID ABL32260 standard; DNA; 15881 BP.

XX

AC ABL32260;  
XX  
DT 26-MAR-2002 (first entry)  
DE  
XX Human immune system associated gene SEQ ID NO: 233.  
KW Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytosatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200200928-A2.  
XX  
PD 03-JAN-2002.  
XX  
XX 02-JUL-2001; 2001WO-EP07537.  
PF  
XX 30-JUN-2000; 2000DE-1032529.  
PR  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-130909/17.  
XX  
PT Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation  
XX  
PS Claim 1; SEQ ID NO 233; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 15881 BP; 4515 A; 204 C; 3588 G; 7574 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 384 Length: 15881  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0  
  
US-09-854-133-587 (1-16) x ABL32260 (1-15881)  
  
QY 10 IleilePheTrpIlePheTrp 16  
| | | | | | | | | | | | | | | |  
Db 11862 ATTATTTTGGATATTGG 11882  
  
RESULT 15  
AAS61069  
ID AAS61069 standard; DNA; 15881 BP.  
XX  
AC AAS61069;  
XX  
DT 29-JAN-2002 (first entry)  
XX  
DE Human gene regulation-associated gene oligonucleotide #24.  
XX

KW Human; Gene regulation-associated gene; severe combined immunodeficiency;  
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;  
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;  
KW renal disease; Preeclampsia; cardiac allograft vascular disease;  
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;  
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;  
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.  
XX  
OS Homo sapiens.  
XX  
PN WO200177375-A2.  
XX  
PD 18-OCT-2001.  
XX  
XX 06-APR-2001; 2001WO-EP03968.  
PF  
XX 06-APR-2000; 2000DE-1019058.  
PR  
PR 07-APR-2000; 2000DE-1019173.  
PR  
PR 30-JUN-2000; 2000DE-1032529.  
PR  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
PI  
XX WPI; 2002-017470/02.  
DR  
XX  
XX  
PT New nucleic acid sequences from chemically modified genes associated  
PT with gene regulation, useful for analysing cytosine methylations for  
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency  
PT disease  
XX  
PS Claim 1; SEQ ID No 25; 26pp; English.  
XX  
CC The invention relates to 224 nucleic acid sequences comprising at least  
CC 18 bases of a chemically pretreated gene associated with gene regulation  
CC selected from 43 known genes (or complementary sequences). The  
CC chemical pretreatment converts cytosine bases unmethylated at the  
CC 5-position to uracil or another base with hybridisation behaviour  
CC dissimilar to cytosine, to enable analysis of cytosine methylations.  
CC The DNA sequences, oligomers (or sets/arrays) and method are  
CC useful in the diagnosis of diseases (or predisposition to diseases)  
CC associated with gene regulation and in therapy of such diseases, by  
CC enabling analysis of the cytosine methylation patterns of such genes,  
CC kits are provided. They are especially useful in diagnosis  
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac  
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,  
CC preeclampsia, graft versus-host disease. The present sequence, is a  
CC sequence included in the sequence data for this specification and is  
CC associated with the human gene regulation-associated genes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 15881 BP; 4515 A; 204 C; 3588 G; 7574 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 384 Length: 15881  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0  
  
US-09-854-133-587 (1-16) x AAS61069 (1-15881)  
  
QY 10 IleilePheTrpIlePheTrp 16  
| | | | | | | | | | | | | | | |  
Db 11862 ATTATTTTGGATATTGG 11882  
  
RESULT 16



ABK311182  
ID ABK311182 standard; DNA; 15881 BP.  
XX  
AC ABK311182;  
XX  
DT 23-APR-2002 (first entry)  
XX  
DE Signal transduction associated gene modified DNA #13.  
XX  
KW Human; signal transduction associated gene; cytosine methylation state;  
KW CpG island; signal transduction associated disease; solid tumour; cancer;  
KW antitumour; cytostatic; mutant; ds.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO200200926-A2.  
XX  
PD 03-JAN-2002.  
XX  
PF 29-JUN-2001; 2001WO-EP07472.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-147896/19.  
XX  
PT Oligonucleotide for diagnosis and therapy of diseases associated with  
PT signal transduction e.g. cancer, comprises chemically modified genomic  
PT sequences of genes associated with signal transduction  
XX  
PS Claim 1; SEQ ID No 25; 24pp; English.  
XX  
CC The present invention relates to chemically modified DNA sequences of  
CC signal transduction associated genes. The DNA sequences are chemically  
CC modified using a solution of bisulphite, hydrogen sulphite or  
CC disulphite. Also disclosed are oligonucleotides and/or pna oligomers  
CC for detecting the cytosine methylation state (CpG islands) of these  
CC genes, and a method for the diagnosis and/or therapy of genetic and  
CC epigenetic parameters of genes associated with signal transduction.  
CC The genomic DNA can be obtained from cells or cellular components which  
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,  
CC cerebral-spinal fluid, tissue embedded in paraffin such as tissue from  
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,  
CC histologic object slides, and all their possible combinations. The  
CC sequences of the invention are useful for the diagnosis and therapy of  
CC diseases associated with signal transduction e.g. solid tumours and  
CC cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA  
CC sequences of different genes associated with signal transduction, or  
CC their complementary sequences.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
SQ Sequence 15881 BP; 4515 A; 204 C; 3588 G; 7574 T; 0 other;

Alignment Scores:  
Pred. No.: 384 Length: 15881  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABK311182 (1-15881)

QY 10 ILeIlePheTrpIlePheTrp 16  
|||||

Db 11862 ATTAATTTTGGATATTTGG 11882

RESULT 17  
ABN79984  
ID ABN79984 standard; DNA; 16633 BP.  
XX  
AC ABN79984;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human chemically modified disease associated gene SEQ ID NO 1.  
XX  
KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
KW antidiabetic; cytostatic; anticonvulsant; ds.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO200200927-A2.  
XX  
PD 03-JAN-2002.  
XX  
PF 02-JUL-2001; 2001WO-EP07536.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-130908/17.  
XX  
PT Novel nucleic acid useful for diagnosis and therapy of diseases  
PT associated with development genes such as diabetes, comprises a  
PT sequence of a segment of chemically pretreated DNA of genes associated  
PT with development  
XX  
PS Claim 1; SEQ ID NO 1; 27pp; English.  
XX  
CC The invention relates to a nucleic acid (I) comprising a sequence at  
CC least 18 bases in length of a segment of chemically pretreated DNA (II)  
CC of genes associated with development selected from 87 genes listed in  
CC the specification such as ACCPN, ADFN, or AFD1 and comprising one of 350  
CC sequences (ABN79984-ABN80333) or their complements. The invention is  
CC useful for the diagnosis or therapy of diseases associated with  
CC development genes, in particular disease related to homeobox containing  
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
CC associated with congenital heart disease, epilepsy, diseases related to  
CC histone deacetylation, Currarino syndrome, diseases related with the  
CC development of the brain and limb girdle muscular dystrophy and dwarfism.  
CC Oligomers specific to each of the genes are useful for detecting the  
CC methylation state of all CpG dinucleotides within the 350 sequences or  
CC (II) and their complementary sequences, as primer oligonucleotides for  
CC the amplification of the 350 sequences, (II) and/or their complements and  
CC as oligomer probes for detecting the cytosine methylation state and/or  
CC single nucleotide polymorphisms (SNPs).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office.

SQ Sequence 16633 BP; 4452 A; 142 C; 3667 G; 8372 T; 0 other;

Alignment Scores:  
Pred. No.: 399 Length: 16633  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABN79984 (1-16633)

QY 9 PheillePheTrpIlePhe 15  
Db 11972 TTTATTATTTTTGGATTTT 11992  
RESULT 18  
AAV42717  
ID AAV42717 standard; DNA; 26 BP.  
XX AC AAV42717;  
XX 14-OCT-1998 (first entry)  
DE 3' PCR primer used to amplify human beta galactosidase cDNA.  
XX Retroviral vector; gene delivery vehicle; expression; PCR primer;  
KW non-immunogenic selectable marker; gene therapy; activation;  
KW human; beta galactosidase; ss.  
XX Synthetic.  
OS Homo sapiens.  
XX WO9830709-A2.  
PN 16-JUL-1998.  
PD 14-JAN-1998; 98WO-US00715.  
XX 13-JAN-1998; 98US-0038339.  
PR 14-JAN-1997; 97US-0035473.  
PR 27-FEB-1997; 97US-0038339.  
XX (CHAD/) CHADA S.  
PA (JOLLY) JOLLY D J.  
PA (MOOR/) MOORE M D.  
XX Chada S, Jolly DJ, Moore MD;  
PI WPI; 1998-399153/34.  
XX Non-immunogenic pro:drug activating enzyme(s) and selectable  
PT marker(s) - are used in gene therapy for the treatment of a wide  
PT variety of disorder(s)  
XX Example 2; Page 46; 121pp; English.  
PS PCR primers AAV42716-17 were used to amplify human beta galactosidase  
CC cDNA (see AAV42728). The amplified product is used as a genetic marker  
CC in the retroviral vector of the invention. The specification describes  
CC a gene delivery vehicle which directs expression of a non-immunogenic  
CC selectable marker or molecule which is capable of activating a  
CC previously inactive compound. Vectors expressing the markers and a  
CC heterologous sequence are useful in gene therapy. The vectors can be used  
CC to deliver a molecule into a target area where it may cause the  
CC activation of a previously inactive substance.  
XX Sequence 26 BP; 7 A; 5 C; 7 G; 7 T; 0 other;  
SQ Alignment Scores:  
Pred. No.: 19.9 Length: 26  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 19 Gaps: 0  
US-09-854-133-587 (1-16) x AAV42717 (1-26)  
QY 6 GlylleAspPheille 11  
Db 4 GGGATCGATTTCATC 21  
RESULT 19

AAC30925/c  
ID AAC30925 standard; cDNA; 246 BP.  
XX AAC30925;  
AC AAC30925;  
XX 06-OCT-2000 (first entry)  
XX Human secreted protein 5' EST, SEQ ID NO: 35000.  
DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping; ss.  
KW Homo sapiens.  
XX EP1033401-A2.  
PN 06-SEP-2000.  
PD 21-FEB-2000; 2000EP-0200610.  
PF 26-FEB-1999; 99US-0122487.  
PR (GEST ) GENSET.  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
PI WPI; 2000-500381/45.  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX Claim 1; SEQ ID 35000; 71pp + CD-ROM; English.  
PS The present sequence is one of a large number of 5' ESTs derived from  
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.  
XX Sequence 246 BP; 74 A; 47 C; 55 G; 70 T; 0 other;  
SQ Alignment Scores:  
Pred. No.: 136 Length: 246  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 21 Gaps: 0  
US-09-854-133-587 (1-16) x AAC30925 (1-246)  
QY 10 IlellePheTrpIlePhe 15  
Db 61 ATTATCTTCTGGATTTTC 44  
RESULT 20  
ABL77073/c  
ID ABL77073 standard; cDNA; 250 BP.  
XX ABL77073;  
AC ABL77073;  
XX 17-MAY-2002 (first entry)  
XX

DE	Human breast cancer cDNA clone 13082.
XX	
KW	Human; ss; breast cancer protein; tumour; cancer; cytostatic;
KW	gene therapy.
XX	
OS	Homo sapiens.
XX	
PN	WO200179286-A2.
XX	
PD	25-OCT-2001.
XX	
PF	12-APR-2001; 2001WO-US12164.
XX	
PR	17-APR-2000; 2000US-0551621.
PR	08-JUN-2000; 2000US-0590751.
PR	22-JUN-2000; 2000US-0604287.
PR	20-JUL-2000; 2000US-0620405.
XX	
PA	(CORI-) CORIXA CORP.
PI	Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepl
XX	
PI	WPI; 2001-611721/70.
DR	
XX	Breast Tumour Proteins and nucleic acids useful for the preven
PT	diagnosis and treatment of breast cancer -
PT	
XX	
PS	Claim 1; Page 175; 297pp; English.
XX	
CC	The invention relates to isolated breast tumour proteins and
CC	nucleic acids that encode them, including immunogenic fragmen
CC	proteins. Also included are expression vectors expressing the
CC	proteins, transformed cells and antibodies raised against the
CC	an antigen presenting cell expressing the protein. The prote
CC	nucleic acids may be used in the prevention, diagnosis and t
CC	diseases associated with inappropriate breast tumour protein
CC	i.e. breast tumours and breast cancer e.g by gene therapy. Th
CC	acids and their complements may also be used as DNA probes i
CC	assays to detect and quantitate the presence of similar nucl
CC	samples, and therefore which patients may be in need of rest
CC	therapy. The proteins, nucleic acids and antibodies may be u
CC	to identify modulators (e.g. antagonists) of breast tumour p
CC	expression and activity. The antibodies and antagonists may
CC	to down regulate expression and activity. The antibodies may
CC	as diagnostic agents for detecting the presence of the prote
CC	samples (e.g. by enzyme linked immunosorbant assay (ELISA))
CC	immuno-purification diagnostic techniques. The present sequen
CC	a cDNA from a breast tumour cDNA library isolated by subtra
CC	hybridisation against a normal breast cDNA library.
XX	
SQ	Sequence 277 BP; 97 A; 44 C; 80 G; 56 T; 0 other;

Alignment Scores:

Pred. No.:	151	Length:	277
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	37.50%	Indels:	0
DB:	22	Gaps:	0

US-09-854-133-587 (1-16) x AAS47046 (1-277)

QY    10 IleIlePheTrpIlePhe 15  
      | | | | | | | | | | | | | | | |  
DB    96 ATTATATTGGATCTTC 113

RESULT 22  
AAFL17616

ID    AAFL17616 standard; cDNA; 277 BP.  
XX  
AC    AAFL17616;  
XX  
DT    13-MAR-2001 (first entry)

XX DE Human breast cancer associated 13082 coding sequence.  
XX KW Human; breast cancer associated gene; vaccine; diagnosis; therapy; ss.  
XX OS Homo sapiens.  
XX PN WO200060076-A2.  
XX PD 12-OCT-2000.  
XX PF 15-FEB-2000; 2000WO-US05308.  
XX PR 02-APR-1999; 99US-0285480.  
XX PR 23-JUN-1999; 99US-0339338.  
XX PR 02-SEP-1999; 99US-0389681.  
XX PR 03-NOV-1999; 99US-0433826.  
XX PA (CORI-) CORIXA CORP.  
XX PI Yuqiu J, Dillon DC, Mitcham JL, Xu J, Harlocker SL;  
XX WPI; 2001-122627/13.  
XX AN An isolated polypeptide useful for the treatment and diagnosis of  
PT tumors e.g. breast cancer comprises at least an immunogenic portion of  
PT a breast tumor protein -  
XX Claim 6; Page 117; 238pp; English.  
XX CC The present invention provides the coding sequences and some protein  
CC sequences of proteins associated with breast cancer in humans. These  
CC sequences can be used in the diagnosis and treatment of cancers,  
CC particularly breast tumours.  
XX SQ Sequence 277 BP; 97 A; 44 C; 80 G; 56 T; 0 other;  
Alignment Scores:  
Pred. No.: 151 Length: 277  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 22 Gaps: 0  
US-09-854-133-587 (1-16) x AAF17616 (1-277)  
QY 10 IleIlePheTrpIlePhe 15  
Db 96 ATTATATTTTGGATCTTC 113  
RESULT 23  
ABK95081  
ID ABK95081 standard; cDNA; 277 BP.  
XX AC ABK95081;  
XX DT 24-SEP-2002 (first entry)  
DE Human breast tumour cDNA 13082.  
XX KW Human; ss; breast cancer; cytostatic; gene therapy; SYN22A12; SYN22A2;  
XX B723P; B726P.  
OS Homo sapiens.  
XX PN US6387697-B1.  
XX PD 14-MAY-2002.  
XX PF 28-DEC-1998; 98US-0222575.  
XX PR 28-DEC-1998; 98US-0222575.

XX PA (CORI-) CORIXA CORP.  
XX PI Yuqiu J, Dillon DC, Mitcham JL, Xu J;  
XX WPI; 2002-478446/51.  
XX PT New nucleic acids encoding breast cancer antigens SYN22A12, SYN22A2,  
PT B723P and B726P, useful for the prevention, diagnosis and treatment of  
XX breast cancer -  
PS Example 2; Column 89-90; 82pp; English.  
XX The invention relates to isolated polynucleotides ((I)-(IV) encoding  
CC breast cancer antigens SYN22A12, SYN22A2, B723P and B726P (4 of 172 cDNA  
CC sequences included in the specification). Also included are isolated  
CC cDNAs comprising sequences with 90 % identity to(I)-IV), an expression  
CC vector comprising the cDNAs, a host cell transformed with the expression  
CC probes, where 1 of the oligonucleotide primers or probes is specific  
CC for (I)-(IV) and isolated cDNAs comprising sequences with 90 %  
CC identity to (I)-(IV), and that is 10 nucleotides in length.  
CC These polynucleotide sequences (I) to (IV) and the SYN22A12, SYN22A2,  
CC B723P and B726P peptides they encode can be used as diagnostic  
CC markers, as models for the development of human therapeutic targets,  
CC aid in the identification of therapeutic proteins, and serve as targets  
CC for the development of human therapeutic agents that may be used for  
CC the treatment of breast cancer. In particular, they may be transfected  
CC (using a vector) into a host cell that may be cultured to express the  
CC breast cancer antigens. These antigens may then be used in the production  
CC of antibodies against SYN22A12, SYN22A2, B723P and B726P, which in turn  
CC may be used as immunoassay reagent and therapeutics for the diagnosis and  
CC treatment of breast cancer. Short 10 nucleotide sequences of (I) to (IV)  
CC may be used as primers and probes in hybridisation and amplification  
CC reactions for the detection of SYN22A12, SYN22A2, B723P and B726P in the  
CC diagnosis of breast cancer. The present sequence is a breast cancer  
CC specific cDNA of the invention.  
XX SQ Sequence 277 BP; 97 A; 44 C; 80 G; 56 T; 0 other;  
Alignment Scores:  
Pred. No.: 151 Length: 277  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 24 Gaps: 0  
US-09-854-133-587 (1-16) x ABK95081 (1-277)  
QY 10 IleIlePheTrpIlePhe 15  
Db 96 ATTATATTTTGGATCTTC 113  
RESULT 24  
ABL87080  
ID ABL87080 standard; cDNA; 283 BP.  
XX AC ABL87080;  
XX DT 17-MAY-2002 (first entry)  
XX DE Human ovarian cancer related cDNA clone SEQ ID NO:10058.  
XX KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200192581-A2.  
XX PD 06-DEC-2001.  
XX PF 29-MAY-2001; 2001WO-US17756.



XX 26-MAY-2000; 2000US-207484P.  
PR (CORI-) CORIXA CORP.  
XX  
PA  
XX  
PI Algate PA, Harlocker SL, Jones R;  
XX WPI; 2002-122075/16.  
DR  
XX  
XX Composition for therapy and diagnosis of ovarian cancer comprising  
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
PT polypeptide, antibody specific to polypeptide or T cell expressing  
PT polypeptide  
XX  
XX Claim 1; SEQ ID 10058; 489pp; English.  
PS  
XX  
CC The present invention describes a composition (I) comprising: carriers  
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour  
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence  
CC (SI) from the 10912 nucleotide sequences as given in ABL77023 to  
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell  
CC population of (II), or antigen presenting cells that express (II).  
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to  
CC (SI) can be used for detecting ovarian cancer in a patient's biological  
CC sample preferably serum or ovarian tissue. The method comprises  
CC contacting a biological sample from a patient with (IV), detecting the  
CC amount of polynucleotide hybridising to (IV) and comparing the amount to  
CC a predetermined cutoff value and thereby detecting ovarian cancer in the  
CC patient, where the amount of polynucleotide hybridising to (IV) is  
CC detected preferably by polymerase chain reaction (PCR). (I) comprising  
CC (III) and/or (II) is useful for stimulating and/or expanding T cells  
CC specific for an ovarian tumour protein comprising contacting T cells  
CC with (III) or (II). (III) is useful in design and preparation of  
CC ribozyme molecules for inhibiting expression of the tumour polypeptides  
CC and proteins in tumour cells; and to isolate a full length gene from a  
CC suitable library e.g., a tumour cDNA library using well known  
CC techniques.  
XX  
SQ Sequence 283 BP; 60 A; 71 C; 58 G; 94 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 153 Length: 283  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 24 Gaps: 0  
  
US-09-854-133-587 (1-16) x ABL87080 (1-283)  
  
QY 9 PheIleIlePheTrpIle 14  
          |||||  
Db 21 TTTATTATTTTGGATA 38  
  
RESULT 25  
AAK77254/C  
ID AAK77254 standard; DNA; 303 BP.  
XX  
AC AAK77254;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32066.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
  
PF 17-JAN-2001; 2001WO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.



PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
PS Claim 1; Page 2881; 11750pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 312 BP; 94 A; 70 C; 53 G; 93 T; 2 other;

Alignment Scores:  
Pred. No.: 167 Length: 312  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x ABV17400 (1-312)

QY 9 PheIleIlePheTrpIle 14  
Db 67 TTCATTATTTTGGATT 50

## RESULT 27

ABV47195/c  
ID ABV47195 standard; cDNA; 339 BP.

XX  
AC ABV47195;

DT 16-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 47186.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.

XX Homo sapiens.

OS WO200160860-A2.

PN 23-AUG-2001.

PD 20-FEB-2001; 2001WO-US05171.

XX 17-FEB-2000; 2000US-183319P.

PR 16-MAR-2000; 2000US-189862P.

PR 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.

PR 18-JUL-2000; 2000US-219007P.

PR 13-DEC-2000; 2000US-255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

PI WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
PS Claim 1; Page 9288; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX  
SQ Sequence 339 BP; 104 A; 81 C; 61 G; 93 T; 0 other;

Alignment Scores:  
Pred. No.: 179 Length: 339  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x ABV47195 (1-339)

QY 9 PheIleIlePheTrpIle 14

Db 118 TTCATTATTTTGGATT 101

## RESULT 28

AAS66468/c  
ID AAS66468 standard; cDNA; 416 BP.

XX  
AC AAS66468;

XX 13-FEB-2002 (first entry)

DT DNA encoding novel human diagnostic protein #2272.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

PD 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

XX P-PSDB; ABG02281.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX

PS Claim 1; SEQ ID No 2272; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations in  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 416 BP; 75 A; 94 C; 61 G; 74 T; 112 other;

Alignment Scores:  
Pred. No.: 213 Length: 416  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x AAS66468 (1-416)

QY 6 GlyIleAspPheIleIle 11  
DB 338 GGCATTGATTTTATCAT 321

RESULT 29  
AAV90289  
ID AAV90289 standard; cDNA; 418 BP.  
XX  
AC AAV90289;  
XX  
DT 15-FEB-1999 (first entry)  
XX  
DE EST clone DK70.  
XX  
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;  
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
KW gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9845436-A2.  
XX  
PD 15-OCT-1998.  
XX  
PF 10-APR-1998; 98WO-US06955.  
XX  
PR 10-APR-1997; 97US-0838821.  
XX  
PA (GEMY) GENETICS INST INC.  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
PI Racie LA, Spaulding V, Treacy M;  
XX  
DR WPI; 1999-070077/06.  
XX  
PT New polynucleotides encoding human secreted proteins - derived from

PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries.  
XX  
PS Claim 1; Page 492-493; 618pp; English.

XX  
CC The present sequence represents a human expressed sequence tag (EST).  
CC The polynucleotide, which is a secreted EST, and the encoded protein  
CC are predicted to have useful biological activities which would make  
CC them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals, although no supporting data is  
CC given. Suggested activities include nutritional activity, immune  
CC stimulating or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.

XX  
SQ Sequence 418 BP; 130 A; 75 C; 99 G; 114 T; 0 other;

Alignment Scores:  
Pred. No.: 214 Length: 418  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 20 Gaps: 0

US-09-854-133-587 (1-16) x AAV90289 (1-418)

QY 9 PheIleIlePheTrpIle 14  
DB 381 TTTATAATATTTGGATA 398

RESULT 30

ABN94582  
ID ABN94582 standard; DNA; 498 BP.

XX  
AC ABN94582;

XX  
DT 13-AUG-2002 (first entry)

XX  
DE Gene #1080 used to diagnose liver cancer.

XX  
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
KW metastatic liver tumour; cytostatic; expression profile; disease state;  
KW disease progression; drug toxicity; drug efficacy; drug metabolism.

XX  
OS Homo sapiens.

XX  
PN WO200229103-A2.

XX  
PD 11-APR-2002.

XX  
PF 02-OCT-2001; 2001WO-US30589.

XX  
PR 02-OCT-2000; 2000US-237054P.

XX  
PA (GENE-) GENE LOGIC INC.

XX  
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

XX  
DR WPI; 2002-426119/45.

XX  
PT Diagnosing and detecting the progression of liver cancer,  
PT hepatocellular carcinoma or metastatic liver tumor in a patient,  
PT involves detecting the level of expression of two or more genes in a  
PT liver tissue sample -

XX  
PS Claim 1; SEQ ID NO 1080; 298pp; English.

XX  
CC The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver



CC tumour in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in ABN93503-ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytotatic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 498 BP; 159 A; 114 C; 93 G; 132 T; 0 other;

Alignment Scores: 249 Length: 498  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 37.50% Indels: 0  
Query Match: 24 Gaps: 0  
DB:

US-09-854-133-587 (1-16) x ABN94582 (1-498)

QY 6 GlyIleAspPheIleIle 11  
|||||  
Db 48 GGCATTGACTTCATTATA 65

RESULT 31  
ABQ35704  
ID ABQ35704 standard; DNA; 530 BP.

XX AC ABQ35704;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 22295.

XX DE Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX PN WO200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-EP10074.

XX PF 01-SEP-2000; 2000DE-1043826.

XX PR 05-SEP-2000; 2000DE-1044543.

XX XX (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX DR WPI; 2002-371829/40.

XX PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA

XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

XX  
SQ Sequence 530 BP; 95 A; 60 C; 146 G; 228 T; 1 other;

Alignment Scores: 263 Length: 530  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 37.50% Indels: 0  
Query Match: 24 Gaps: 0  
DB:

US-09-854-133-587 (1-16) x ABQ35704 (1-530)

QY 7 IleAspPheIleIlePhe 12  
|||||  
Db 190 ATAGATTTTATTATTTT 207

RESULT 32  
ABQ35705/C  
ID ABQ35705 standard; DNA; 530 BP.

XX AC ABQ35705;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 22296.

XX DE Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX PN WO200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-EP10074.

XX PF 01-SEP-2000; 2000DE-1043826.

XX PR 05-SEP-2000; 2000DE-1044543.

XX XX (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX DR WPI; 2002-371829/40.

XX PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA

XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.

CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised from the  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

SQ Sequence 530 BP; 228 A; 146 C; 60 G; 95 T; 1 other;

Alignment Scores:

Pred. No.: 263 Length: 530  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABQ35705 (1-530)

QY 7 IleAspPheillePhe 12

Db 341 ATAGATTATTATATTT 324

RESULT 33

AAH11384  
ID AAH11384 standard; cDNA; 548 BP.

XX AC AAH11384;

XX 26-JUN-2001 (first entry)

DE Human cDNA clone (3'-primer) SEQ ID NO:8219.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
PS Claim 3; SEQ ID 8219; 2537pp + CD ROM; English.

XX

CC The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

SQ Sequence 548 BP; 115 A; 161 C; 133 G; 130 T; 9 other;

Alignment Scores:

Pred. No.: 270 Length: 548  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 22 Gaps: 0

US-09-854-133-587 (1-16) x AAH11384 (1-548)

QY 1 PheGlnAlaAsnCysGly 6

Db 296 TTTCAGGCCAATTGCGGA 313

RESULT 34

ABQ45838

ID ABQ45838 standard; DNA; 553 BP.

XX AC ABQ45838;

XX 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 32429.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
PI WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

XX Sequence 553 BP; 105 A; 65 C; 190 G; 193 T; 0 other;

Alignment Scores:  
Pred. No.: 272 Length: 553  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABQ45838 (1-553)

QY 5 CysGlyIleAspPheille 10

DB 251 TCGCGAATCGATTATT 268

RESULT 35

ABQ45839/C

ID ABQ45839 standard; DNA; 553 BP.

XX  
AC ABQ45839;

XX  
DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 32430.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.

XX Homo sapiens.

OS WO200218632-A2.

XX  
PN 07-MAR-2002.

XX  
PD 01-SEP-2001; 2001WO-EP10074.

XX  
PF 01-SEP-2000; 2000DE-1043826.

XX  
PR 05-SEP-2000; 2000DE-1044543.

XX  
XX (EPIG-) EPIGENOMICS AG.

XX  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX  
XX WPI; 2002-371829/40.

XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

XX Sequence 553 BP; 193 A; 190 C; 65 G; 105 T; 0 other;

Alignment Scores:  
Pred. No.: 272 Length: 553  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABQ45839 (1-553)

QY 5 CysGlyIleAspPheille 10

DB 303 TCGCGAATCGATTATT 286

RESULT 36

ABQ44134

ID ABQ44134 standard; DNA; 559 BP.

XX  
AC ABQ44134;

XX  
DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 30725.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.

XX Homo sapiens.

OS WO200218632-A2.

XX  
PN 07-MAR-2002.

XX  
PD 01-SEP-2001; 2001WO-EP10074.

XX  
PF 01-SEP-2000; 2000DE-1043826.

XX  
PR 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

XX  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;



XX DR WPI; 2002-371829/40.

XX PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
XX amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

XX SQ Sequence 559 BP; 67 A; 68 C; 216 G; 208 T; 0 other;

Alignment Scores:

Pred. No.: 275 Length: 559  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABQ44134 (1-559)

QY 8 AspPheIleIlePheTrp 13  
Db 517 GATTTATTATTTTGG 534

RESULT 37  
ABQ44135/C

ID ABQ44135 standard; DNA; 559 BP.

XX AC ABQ44135;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 30726.

XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX PN WO200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-EP10074.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PR 05-SEP-2000; 2000DE-1044543.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX DR WPI; 2002-371829/40.

XX PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
XX amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

XX SQ Sequence 559 BP; 208 A; 216 C; 68 G; 67 T; 0 other;

Alignment Scores:

Pred. No.: 275 Length: 559  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABQ44135 (1-559)

QY 8 AspPheIleIlePheTrp 13  
Db 43 GATTTATTATTTTGG 26

RESULT 38  
AAD05092/C

ID AAD05092 standard; CDNA; 572 BP.

XX AC AAD05092;

XX DT 17-JUL-2001 (first entry)

XX DE Human secreted protein-encoding gene 9 cDNA clone HDPAQ55, SEQ ID NO:50.

XX KW Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder;  
KW endocrine disorder; infection; wound healing; gene therapy;  
KW cell culture; chemotaxis; food additive;  
XX binding partner identification; ss.

OS Homo sapiens.

XX Key Location/Qualifiers  
FH



FT 17..178  
FT /\*tag= a  
FT /product= "Human secreted protein"  
FT 17..85  
FT /\*tag= b  
FT 86..175  
FT /\*tag= c  
FT /product= "Mature human secreted protein"  
FT

XX  
PN WO200134768-A2.  
XX 17-MAY-2001.  
XX 01-NOV-2000; 2000WO-US30039.  
XX 09-NOV-1999; 99US-0164344.  
PR 07-APR-2000; 2000US-0195296.  
PR 27-JUL-2000; 2000US-0221367.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Olsen HS, Komatsoulis G, Duan DR, Ebner R, Ruben SM;  
PI  
XX  
DR WPI; 2001-308780/32.  
DR P-PSDB; AAE01203.  
XX

Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition -  
Claim 1; Page 430; 474pp; English.  
AAD05053-AAD05106 represent cDNAs corresponding to 15 human secreted protein genes, and AAE01164-AAE01217 represent the proteins they encode. AAE01218-AAE01226 represent human secreted protein fragments or variants. The secreted proteins and their genes are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 15 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angiogenic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplant, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein-encoding cDNA of the invention.

SQ Sequence 572 BP; 162 A; 98 C; 125 G; 179 T; 8 other;

Alignment Scores:  
Pred. No.: 280 Length: 572  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 22 Gaps: 0

US-09-854-133-587 (1-16) x AAD05092 (1-572)

QY 7 IleAspPheIleIlePhe 12  
Db 133 ATTGATTTTATCATATTC 116  
RESULT 39  
AAL34901  
ID AAL34901 standard; cDNA; 579 BP.  
XX  
AC AAL34901;  
XX  
DT 08-JAN-2002 (first entry)  
XX  
DE Human musculoskeletal system related polynucleotide SEQ ID NO 243.  
XX  
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200155367-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01338.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.

ID AAS85327 standard; cDNA: 582 BP.

XX AAS85327;  
AC 13-FEB-2002 (first entry)  
XX DNA encoding novel human diagnostic protein #21131.  
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
KW Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
PI WPI; 2001-639362/73.  
XX P-PSDB; ABG21140.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX Claim 1; SEQ ID No 21131; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 582 BP; 170 A; 132 C; 158 G; 122 T; 0 other;  
SQ  
Alignment Scores: 284 Length: 582  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 37.50% Gaps: 0  
DB: 23  
US-09-854-133-587 (1-16) x AAS85327 (1-582)  
QY 6 GlyIleAspPheIleIle 11  
DB 136 GGTATTGACTTCATAA 153  
RESULT 41  
AAC08897/C  
ID AAC08897 standard; cDNA; 584 BP.  
XX AAC08897;  
AC 06-OCT-2000 (first entry)  
XX Human secreted protein 5' EST, SEQ ID NO: 12972.  
DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX Homo sapiens.  
OS EP1033401-A2.  
XX 06-SEP-2000.  
XX 21-FEB-2000; 2000EP-0200610.  
PF 26-FEB-1999; 99US-0122487.  
XX (GEST ) GENSET.  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
PI WPI; 2000-500381/45.  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX Claim 1; SEQ ID 12972; 71pp + CD-ROM; English.  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.  
XX Sequence 584 BP; 142 A; 112 C; 132 G; 198 T; 0 other;  
SQ  
Alignment Scores: 285 Length: 584  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 37.50% Gaps: 0  
DB: 21  
US-09-854-133-587 (1-16) x AAC08897 (1-584)  
QY 6 GlyIleAspPheIleIle 11  
DB 584 GGCATTGACTTCATTATA 567  
RESULT 42  
ABA62551  
ID ABA62551 standard; DNA; 600 BP.  
XX ABA62551;  
AC 01-FEB-2002 (first entry)  
XX

XX DE Human foetal liver single exon nucleic acid probe #10856.  
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX OS Homo sapiens.  
PN WO200157277-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00669.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX DR WPI; 2001-483447/52.  
XX PT Human genome-derived single exon nucleic acid probes useful for  
XX PT analyzing gene expression in human fetal liver -  
PS Claim 1; SEQ ID NO 10856; 639pp + sequence listing; English.  
XX CC The invention relates to a single exon nucleic acid probe for  
XX CC measuring human gene expression in a sample derived from human foetal  
XX CC liver. The single exon nucleic acid probes may be used for predicting,  
XX CC measuring and displaying gene expression in samples derived from human  
XX CC foetal liver. The present sequence is a single exon nucleic acid  
XX CC probe of the invention.  
XX CC Note: The sequence data for this patent did not form part of the  
XX CC printed specification, but was obtained in electronic format directly  
XX CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 600 BP; 165 A; 132 C; 122 G; 181 T; 0 other;

Alignment Scores:  
Pred. No.: 292 Length: 600  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 22 Gaps: 0

US-09-854-133-587 (1-16) x ABA62551 (1-600)

QY 6 GlyIleAspPheIleIle 11  
Db 489 GGCATTGACTTCATCAT 506

RESULT 43  
ABA29879  
ID ABA29879 standard; DNA; 600 BP.  
XX ABA29879;  
AC ABA29879;  
XX 23-JAN-2002 (first entry)  
DT  
XX Probe #8345 for gene expression analysis in human heart cell sample.  
DE Human; gene expression; heart; microarray; vascular system; probe;  
XX cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease; ss.  
XX OS Homo sapiens.

XX PN WO200157274-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00666.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX DR WPI; 2001-488899/53.  
XX PT Single exon nucleic acid probes for analyzing gene expression in human  
XX PT hearts -  
PS Claim 1; SEQ ID No 8345; 530pp; English.  
XX CC The present invention relates to single exon nucleic acid probes for  
XX CC measuring human gene expression in a sample derived from human heart. The  
XX CC present sequence is one such probe. The probes may be used for  
XX CC predicting, measuring and displaying gene expression in samples derived  
XX CC from the human heart via microarrays. By measuring gene expression, the  
XX CC probes are useful for predicting, diagnosing, grading, staging,  
XX CC monitoring and prognosing diseases of the human heart and vascular system  
XX CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
XX CC congenital heart disease.  
XX CC Note: The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 600 BP; 165 A; 132 C; 122 G; 181 T; 0 other;

Alignment Scores:  
Pred. No.: 292 Length: 600  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 22 Gaps: 0

US-09-854-133-587 (1-16) x ABA29879 (1-600)

QY 6 GlyIleAspPheIleIle 11  
Db 489 GGCATTGACTTCATCAT 506

RESULT 44  
AAK10907  
ID AAK10907 standard; DNA; 600 BP.  
XX AAK10907;  
AC AAK10907;  
XX 05-NOV-2001 (first entry)  
DT  
XX Human brain expressed single exon probe SEQ ID NO: 10898.  
DE Human; brain expressed exon; gene expression analysis; probe;  
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.  
XX OS Homo sapiens.  
XX PN WO200157275-A2.  
XX



```
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
PT
XX
XX Example 4; SEQ ID NO: 10898; 650pp + Sequence Listing; English.
PS
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
XX Sequence 600 BP; 165 A; 132 C; 122 G; 181 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 292 Length: 600
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 22 Gaps: 0
US-09-854-133-587 (1-16) x AAK10907 (1-600)
QY 6 GlyIleAspPheIleIle 11
Db 489 GGCATTGACTTCATCATT 506
RESULT 45
AAK36765
ID AAK36765 standard; DNA; 600 BP.
XX
AC AAK36765;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 11322.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US00668.
PF
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
```

```
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
PT
XX
XX Example 4; SEQ ID NO: 11322; 658pp + Sequence Listing; English.
PS
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
XX Sequence 600 BP; 165 A; 132 C; 122 G; 181 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 292 Length: 600
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 22 Gaps: 0
US-09-854-133-587 (1-16) x AAK36765 (1-600)
QY 6 GlyIleAspPheIleIle 11
Db 489 GGCATTGACTTCATCATT 506
Search completed: May 11, 2003, 17:16:11
Job time : 45.2655 secs
```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 16:16:15 ; Search time 8.92035 Seconds  
(without alignments)  
550.071 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FOANGGIDFIIFWIFW 16

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 873169

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 60 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_174133-728/app\_query.fasta\_1.462  
-DB=Issued\_Patents\_NA -QFWT=fastap -SUFFIX=oligo.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=60 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=45 -MODE=LOCAL  
-OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn\_1\_1\_47 @runat\_05052003\_174133-728 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMMap -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	6	37.5	277	4	US-09-222-575-101
2	6	37.5	1352	4	US-08-817-913-15
3	6	37.5	1470	4	US-09-134-001C-2404
4	6	37.5	1485	4	US-09-134-001C-1339
5	6	37.5	1734	4	US-08-817-913-16
6	6	37.5	1920	4	US-08-817-913-17
7	6	37.5	1969	1	US-07-721-761A-35
8	6	37.5	1969	1	US-07-978-687-35
9	6	37.5	1969	4	US-08-926-522-17
10	6	37.5	1969	5	PCT-US91-05801-35
11	6	37.5	2371	2	US-09-020-466-1
12	6	37.5	2371	3	US-09-192-659-1

13	6	37.5	2453	4	US-08-961-527-316	Sequence 316, Appl
14	6	37.5	2682	4	US-09-020-465-1	Sequence 1, Appli
15	6	37.5	3210	3	US-08-613-009A-6	Sequence 6, Appli
16	6	37.5	3210	4	US-08-778-570B-6	Sequence 6, Appli
17	6	37.5	3210	4	US-09-059-584-6	Sequence 6, Appli
18	6	37.5	3334	4	US-09-255-984-1	Sequence 1, Appli
19	6	37.5	3539	4	US-09-508-542-16	Sequence 16, Appli
20	6	37.5	3660	3	US-08-613-009A-5	Sequence 5, Appli
21	6	37.5	3660	4	US-08-778-570B-5	Sequence 5, Appli
22	6	37.5	3660	4	US-09-059-584-5	Sequence 5, Appli
23	6	37.5	4711	1	US-08-414-926A-1	Sequence 1, Appli
24	6	37.5	4711	2	US-08-926-922-1	Sequence 1, Appli
25	6	37.5	4711	3	US-09-253-682-1	Sequence 1, Appli
26	6	37.5	4711	4	US-09-527-657-1	Sequence 1, Appli
27	6	37.5	5064	4	US-08-936-107A-8	Sequence 8, Appli
28	6	37.5	8654	4	US-08-961-527-98	Sequence 98, Appli
29	6	37.5	10357	4	US-08-961-527-191	Sequence 191, Appl
30	6	37.5	20303	1	US-08-370-975B-6	Sequence 6, Appli
31	6	37.5	26764	1	US-08-370-975B-1	Sequence 1, Appli
32	5	31.2	24	4	US-09-341-444A-28	Sequence 28, Appli
33	5	31.2	48	4	US-08-985-659-2	Sequence 2, Appli
34	5	31.2	66	3	US-09-046-247-138	Sequence 138, App
35	5	31.2	117	1	US-08-450-834-4	Sequence 4, Appli
36	5	31.2	222	4	US-09-134-001C-548	Sequence 548, App
37	5	31.2	225	4	US-09-134-001C-2829	Sequence 2829, Ap
38	5	31.2	258	1	US-08-410-804-6	Sequence 6, Appli
39	5	31.2	258	1	US-08-259-514-6	Sequence 6, Appli
40	5	31.2	258	2	US-08-858-311-6	Sequence 6, Appli
41	5	31.2	313	4	US-09-438-906-15	Sequence 15, Appl
42	5	31.2	390	4	US-09-134-001C-2354	Sequence 2354, Ap
43	5	31.2	405	4	US-09-134-001C-2395	Sequence 2395, Ap
44	5	31.2	409	1	US-08-469-667-21	Sequence 21, Appl
45	5	31.2	409	4	US-09-224-110-21	Sequence 21, Appl
46	5	31.2	409	5	PCT-US95-07289-21	Sequence 21, Appl
47	5	31.2	439	4	US-09-221-017B-162	Sequence 162, App
48	5	31.2	520	4	US-08-961-527-344	Sequence 344, App
49	5	31.2	583	4	US-09-221-017B-682	Sequence 682, App
50	5	31.2	585	4	US-08-976-259-45	Sequence 45, Appl
51	5	31.2	595	4	US-09-385-982-452	Sequence 452, App
52	5	31.2	601	4	US-09-814-951A-13	Sequence 13, Appl
53	5	31.2	611	4	US-09-376-113-4	Sequence 4, Appli
54	5	31.2	624	4	US-09-221-017B-43	Sequence 43, Appli
55	5	31.2	631	4	US-09-376-113-6	Sequence 6, Appli
56	5	31.2	642	2	US-08-480-753-5	Sequence 5, Appli
57	5	31.2	642	3	US-09-041-889-10	Sequence 10, Appl
58	5	31.2	642	3	US-08-837-058-10	Sequence 10, Appl
59	5	31.2	660	2	US-08-401-068-5	Sequence 5, Appli
60	5	31.2	660	2	US-08-846-338-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1  
US-09-222-575-101.  
; Sequence 101, Application US/09222575  
; Patent No. 6387697  
; GENERAL INFORMATION:  
; APPLICANT: Yuqiu, Jiang  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Xu, Jiangchun  
; TITLE OF INVENTION: Compositions for the Treatment and Diagnosis of Breast Can  
; TITLE OF INVENTION: and Methods for Their Use  
; FILE REFERENCE: 210121.470  
; CURRENT APPLICATION NUMBER: US/09/222,575  
; CURRENT FILING DATE: 1998-12-28  
; NUMBER OF SEQ ID NOS: 174  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 101  
; LENGTH: 277  
; TYPE: DNA  
; ORGANISM: Human  
US-09-222-575-101

277  
60000

ENCE FROM POTATO

on & Bear  
e 16th Floor

96

001APC

```
length: 135
atches: 6
nservative: 0
smatches: 0
dels: 0
ops: 0
```

15 (1-1352)

4,001C  
964  
779

```

Length:      1470
Matches:      6
Conservative: 0
Mismatches:  0
Indels:      0
Gaps:        0

```

LC-2404 (1-1470)

C

64  
79

length:	1485
tches:	6
nservative:	0
smatches:	0
del:	0
ss:	0

ps: 0  
-1339 (1-1485)



```

APPLICANT: Lund, Marianne
APPLICANT: Okkels, Finn
APPLICANT: Kreiberg, Jette
TITLE OF INVENTION: PROMOTER SEQUENCE FROM POTATO
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,913
FILING DATE: 15-SEP-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/02196
FILING DATE: 06-JUN-1995
APPLICATION NUMBER: GB941286.7
FILING DATE: 21-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: DYOU10.001APC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714-760-0404
TELEFAX: 714-760-9502
TELEX:
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1920 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-817-913-17
Alignment Scores:
Pred. No.: 108
Score: 6.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 37.50%
DB: 4
Length:
Matches:
Conservative:
Mismatch:
Indels:
Gaps:
US-09-854-133-587 (1-16) x US-08-817-913-17 (1-1920)
QY 7 IleAspPheIleIlePhe 12
Db 240 ATAGACTTCATAATTTT 257
RESULT 7
US-07-721-761A-35
; Sequence 35, Application US/07721761A
; Patent No. 5475099
; GENERAL INFORMATION:
; APPLICANT: Vic. C. Knauf
; APPLICANT: Gregory A. Thompson
; TITLE OF INVENTION: Plant Fatty Acid Synthases
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616

```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.7
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/721,761A
; FILING DATE: 19910626
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/568,493
; FILING DATE: 15-AUGUST-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; REFERENCE/DOCKET NUMBER: CGNE 76-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1969 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-07-721-761A-35

Alignment Scores:
Pred. No.: 110 Length: 1969
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 1 Gaps: 0

US-09-854-133-587 (1-16) x US-07-721-761A-35 (1-1969)
QY 9 PheillelPheTrpIle 14
Db 160 TTTATCATCTTCTGGATC 177

RESULT 8
US-07-978-687-35
; Sequence 35, Application US/07978687
; Patent No. 5510255
; GENERAL INFORMATION:
; APPLICANT: Vic C. Knauf
; APPLICANT: Gregory A. Thompson
; TITLE OF INVENTION: Plant Fatty Acid Synthases
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.7
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/978,687
; FILING DATE: FEBRUARY 1, 1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/05801
; FILING DATE: 15-AUGUST-1991
; PRIOR APPLICATION DATA:
```

```
;
; APPLICATION NUMBER: 07/568,493
; FILING DATE: 15-AUGUST-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/721,761
; FILING DATE: 26-JUNE-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; REFERENCE/DOCKET NUMBER: CGNE 76-2 WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1969 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-07-978-687-35

Alignment Scores:
Pred. No.: 110 Length: 1969
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 1 Gaps: 0

US-09-854-133-587 (1-16) x US-07-978-687-35 (1-1969)
QY 9 PheillelPheTrpIle 14
Db 160 TTTATCATCTTCTGGATC 177

RESULT 9
US-08-926-522-17
; Sequence 17, Application US/08926522
; Patent No. 6426447
; GENERAL INFORMATION:
; APPLICANT: Vic C. Knauf
; APPLICANT: Gregory A. Thompson
; TITLE OF INVENTION: PLANT SEED OILS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.1
; SOFTWARE: Microsoft Word 5.1(a)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/458,173
; FILING DATE: 2-June-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE DES
```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1969 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-08-926-522-17

```

Alignment Scores:	110	Length:	1969
Pred. No.:	6.00	Matches:	6
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	37.50%	Gaps:	0
DB:	4		

US-09-854-133-587 (1-16) X US-08-926-522-17 (1-1969)

Qy 9 pheillelephetrpile 14  
 |||||  
 Db 160 TTTATCATCTTCTGGATC 177

RESULT 10  
PCT-US91-05801-35  
; Sequence 35, Application PC/TUS9105801  
; GENERAL INFORMATION:  
; APPLICANT: Vic. C. Knauf  
; APPLICANT: Gregory A. Thompson  
; TITLE OF INVENTION: plant Fatty Acid Synthases  
; NUMBER OF SEQUENCES:  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Calgene, Inc.  
; STREET: 1920 Fifth Street  
; CITY: Davis  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 95616

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.7
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/05801
;

```

APPLICATION NUMBER: 10270554/05004  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA: 07/568,493  
 APPLICATION NUMBER: 15-AUGUST-1990  
 FILING DATE:  
 PRIOR APPLICATION DATA: 07/721,761  
 APPLICATION NUMBER: 26-JUNE-1991  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Elizabeth Lassen  
 REGISTRATION NUMBER: 31,845  
 NAME: Donna E. Scherer  
 REGISTRATION NUMBER: 34,719  
 REFERENCE/DOCKET NUMBER: CGNE 76-2 WO

Alignment Scores:		
Pred. No.:	110	Length:
Score:	6.00	Matches:
Percent Similarity:	100.00%	Conservative:
Best Local Similarity:	100.00%	Mismatches:
Query Match:	37.50%	Indels:
DR:	5	Gaps:
		1969
		6
		0
		0
		0
		0

US-09-854-133-587 (1-16) x PCT-US91-05801-35 (1-1969)

Qy 9 PheIleIlePheTrpIle 14  
|||  
Db 160 TTTATCATCTTCTGGATC 177

RESULT 11  
US-09-020-466-1  
; Sequence 1, Application US/09020466  
; Patent No. 5879908  
; GENERAL INFORMATION:  
; APPLICANT: LAPING, NICHOLAS  
; APPLICANT: OLSON, BARBARA  
; APPLICANT: ZHU, YUAN  
; TITLE OF INVENTION: CRFG-1a, a target and marker  
; TITLE OF INVENTION: for chronic renal failure  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: RATNER & PRESTIA  
; STREET: P.O. BOX 980  
; CITY: VALLEY FORGE  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19482

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/020,466  
FILING DATE: 09-FEB-1998

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/045,203  
FILING DATE: 30-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH-70009-2  
TELECOMMUNICATION INFORMATION:

```

; TELECOMMUNICATION INFORMATION
;
; TELEPHONE: 610-407-0700
;
; TELEFAX: 610-407-0701
;
; TELEX: 846169
;
; INFORMATION FOR SEQ ID NO: 1:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 2371 base pairs
;
; TYPE: nucleic acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: CDNA
;
; US-09-020-466-1

```

Alignment Scores:		
Pred. No.:	130	Length: 2371
Score:	6.00	Matches: 6
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	37.50%	Indels: 0
DR:	2	Gaps: 0

US-09-854-133-587 (1-16) x US-09-020-466-1 (1-2371)

0y 10 IleIlePheTrpIlePhe 15

```
Db 1226 ATTATATTGGATCTTC 1243
RESULT 12
US-09-192-659-1
; Sequence 1, Application US/09192659
; Patent No. 6127522
; GENERAL INFORMATION:
; APPLICANT: LAPING, NICHOLAS J.
; APPLICANT: OLSON, BARBARA
; APPLICANT: ZHU, YUAN
; TITLE OF INVENTION: CRFG-1a, A TARGET AND MARKER
; TITLE OF INVENTION: FOR CHRONIC RENAL FAILURE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09192,659
; FILING DATE: 16-NOV-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/020,466
; FILING DATE: 09-FEB-1998
; APPLICATION NUMBER: 60/045,203
; FILING DATE: 30-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70009-3
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0700
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2371 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-09-192-659-1
Alignment Scores:
Pred. No.: 130 Length: 2371
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 3 Gaps: 0
US-09-854-133-587 (1-16) x US-09-192-659-1 (1-2371)
QY 10 IleIlePheTrpIlePhe 15
Db 1226 ATTATATTGGATCTTC 1243
RESULT 13
US-09-961-527-316
; Sequence 316, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
```

```
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 316:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2453 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-316
Alignment Scores:
Pred. No.: 133 Length: 2453
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 4 Gaps: 0
US-09-854-133-587 (1-16) x US-08-961-527-316 (1-2453)
QY 9 PheIleIlePheTrpIle 14
Db 940 TTTATTATCTTTTGGATA 957
RESULT 14
US-09-020-465-1
; Sequence 1, Application US/09020465
; Patent No. 6255471
; GENERAL INFORMATION:
; APPLICANT: LAPING, NICHOLAS
; APPLICANT: OLSON, BARBARA
; APPLICANT: ZHU, YUAN
; TITLE OF INVENTION: CRFG-1b, a target and marker
; TITLE OF INVENTION: for chronic renal failure
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
```



APPLICATION NUMBER: US/09/020,465  
FILING DATE: 09-FEB-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/045,203  
FILING DATE: 30-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH-70009-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2682 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-020-465-1

Alignment Scores:  
Pred. No.: 144 Length: 2682  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-020-465-1 (1-2682)

QY 10 IleIlePheTrpIlePhe 15  
|||||  
Db 1537 ATTATATTGGATCTTC 1554

## RESULT 15

US-08-613-009A-6/C  
Sequence 6, Application US/08613009A  
Patent No. 6090576  
GENERAL INFORMATION:  
APPLICANT: Myers, Lisa E  
APPLICANT: Schryvers, Anthony B  
APPLICANT: Harkness, Robin E  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Du, Run-Pan  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/613,009A  
FILING DATE: 08-MAR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-542  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3210 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-613-009A-6

Alignment Scores:  
Pred. No.: 169 Length: 3210  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 3 Gaps: 0

US-09-854-133-587 (1-16) x US-08-613-009A-6 (1-3210)

QY 11 IlePheTrpIlePheTrp 16  
|||||  
Db 28 ATTTTGGATTTTGG 11

## RESULT 16

US-08-778-570B-6/C  
Sequence 6, Application US/08778570B  
Patent No. 6437096  
GENERAL INFORMATION:  
APPLICANT: Myers, Lisa E  
APPLICANT: Schryvers, Anthony B  
APPLICANT: Harkness, Robin E  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Du, Run-Pan  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/778,570B  
FILING DATE: 03-JAN-1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-664  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3210 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-778-570B-6

Alignment Scores:  
Pred. No.: 169 Length: 3210  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 37.50% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-08-778-570B-6 (1-3210)

QY 11 IlePheTrpIlePheTrp 16  
|||||  
Db 28 ATTTTGGATTTTTGG 11

RESULT 17

US-09-059-584-6/c  
; Sequence 6, Application US/09059584  
; Patent No. 6440701  
; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/059,584  
; FILING DATE: 14-APR-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/778,570  
; FILING DATE: 03-JAN-1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-794  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3210 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-059-584-6

Alignment Scores:  
Pred. No.: 169 Length: 3210  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-059-584-6 (1-3210)

QY 11 IlePheTrpIlePheTrp 16  
|||||  
Db 28 ATTTTGGATTTTTGG 11

RESULT 18

US-09-255-984-1/c  
; Sequence 1, Application US/09255984  
; Patent No. 6296851  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Richard L.  
; TITLE OF INVENTION: No. 6296851el Compounds  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert, Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2793  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/255,984  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/923,485  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dickinson, Q. Todd  
; REGISTRATION NUMBER: 28,354  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215/994-2252  
; TELEFAX: 215/994-2222  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3334 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-255-984-1

Alignment Scores:  
Pred. No.: 174 Length: 3334  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-255-984-1 (1-3334)

QY 9 PheIleIlePheTrpIle 14  
|||||  
Db 1250 TTTATCATCTTCTGGATA 1233

RESULT 19

US-09-508-542-16  
; Sequence 16, Application US/09508542  
; Patent No. 6339174  
; GENERAL INFORMATION:  
; APPLICANT: STRAUSS, ANDREAS  
; APPLICANT: THUMM, GUNTHER  
; APPLICANT: POHLNER, JOHANNES  
; APPLICANT: GOTZ, FRIEDRICH  
; TITLE OF INVENTION: METHOD FOR IDENTIFYING A NUCLEIC ACID  
; FILE REFERENCE: 10496/P65266US0  
; CURRENT APPLICATION NUMBER: US/09/508,542  
; CURRENT FILING DATE: 2000-05-16  
; PRIOR APPLICATION NUMBER: PCT/EP98/06136  
; PRIOR FILING DATE: 1998-09-26  
; PRIOR APPLICATION NUMBER: 97 116 841.4  
; PRIOR FILING DATE: 1997-09-27

; PRIOR APPLICATION NUMBER: 97 118 755.4  
; PRIOR FILING DATE: 1997-10-29  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 3539  
; TYPE: DNA  
; ORGANISM: Staphylococcus carnosus  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (1)..(3539)  
; OTHER INFORMATION: "n" represents a, t, c, g, unknown or other  
US-09-508-542-16

Alignment Scores: 183 Length: 3539  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 37.50% Gaps: 0  
DB: 4

US-09-854-133-587 (1-16) x US-09-508-542-16 (1-3539)

QY 6 GlyIleAspPheIleIle 11  
|||||  
Db 565 GGGATTGACTTTATTATC 582

## RESULT 20

US-08-613-009A-5/c  
; Sequence 5, Application US/08613009A  
; Patent No. 6090576

## GENERAL INFORMATION:

; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/613,009A  
; FILING DATE: 08-MAR-1996  
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-542  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3660 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-613-009A-5

Alignment Scores: 189 Length: 3660  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 37.50% Gaps: 0  
DB: 3

US-09-854-133-587 (1-16) x US-08-613-009A-5 (1-3660)

QY 11 IlePheTrpIlePheTrp 16  
|||||  
Db 295 ATTTTGGATTTTGG 278

## RESULT 21

US-08-778-570B-5/c  
; Sequence 5, Application US/08778570B  
; Patent No. 6437096

## GENERAL INFORMATION:

; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/778,570B  
; FILING DATE: 03-JAN-1997  
; CLASSIFICATION: 536

## ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-664  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3660 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-778-570B-5

Alignment Scores: 189 Length: 3660  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 37.50% Gaps: 0  
DB: 4

US-09-854-133-587 (1-16) x US-08-778-570B-5 (1-3660)

QY 11 IlePheTrpIlePheTrp 16  
|||||  
Db 295 ATTTTGGATTTTGG 278

## RESULT 22

US-09-059-584-5/c  
; Sequence 5, Application US/09059584  
; Patent No. 6440701  
; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/059,584  
; FILING DATE: 14-APR-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/778,570  
; FILING DATE: 03-JAN-1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-794  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3660 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-059-584-5

Alignment Scores:  
Pred. No.: 189 Length: 3660  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-059-584-5 (1-3660)

QY 11 IlePheTrpIlePheTrp 16  
| | | | | | | | | | | | | | | | | |  
Db 295 ATTTTGGATTTTTGG 278

## RESULT 23

US-08-414-926A-1/c  
; Sequence 1, Application US/08414926A  
; Patent No. 5721354  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; APPLICANT: Cha, Tai-An  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum  
; STREET: 5 Palo Alto Square  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-2155  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/414,926A  
; FILING DATE: March 31, 1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: AVIR-011/COUS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-494-7622  
; TELEFAX: 415-857-0663  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4711 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Human CMV  
; STRAIN: Towne  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: complement (845..1321)  
; OTHER INFORMATION: /product= "UL147"  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: complement (1368..1721)  
; OTHER INFORMATION: /product= "UL152"  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: complement (2504..3337)  
; OTHER INFORMATION: /product= "UL153"  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: complement (3515..4711)  
; OTHER INFORMATION: /product= "UL154"  
; US-08-414-926A-1

Alignment Scores:  
Pred. No.: 235 Length: 4711  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 1 Gaps: 0

US-09-854-133-587 (1-16) x US-08-414-926A-1 (1-4711)

QY 9 PheIleIlePheTrpIle 14  
| | | | | | | | | | | | | | | | | |  
Db 4654 TTTATTATTTTGGATT 4637

## RESULT 24

US-08-926-922-1/c  
; Sequence 1, Application US/08926922  
; Patent No. 5925751  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; APPLICANT: Cha, Tai-An



```

; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cserr Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,922
; FILING DATE: September 10, 1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4711 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Human CMV
; STRAIN: Towne
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (845..1321)
; OTHER INFORMATION: /product= "UL147"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (1368..1721)
; OTHER INFORMATION: /product= "UL152"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (2504..3337)
; OTHER INFORMATION: /product= "UL153"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (3515..4711)
; OTHER INFORMATION: /product= "UL154"
; US-08-926-922-1
Alignment Scores:
Pred. No.: 235 Length: 4711
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 2 Gaps: 0
US-09-854-133-587 (1-16) x US-08-926-922-1 (1-4711)
QY 9 PheilleillePheTrpIle 14
Db 4654 TTTATTATTTTGGATT 4637
RESULT 25
US-09-253-682-1/c
; Sequence 1, Application US/09253682
; Patent No. 6040170
```

```

; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cserr Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/253,682
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/926,922
; FILING DATE: September 10, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4711 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Human CMV
; STRAIN: Towne
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (845..1321)
; OTHER INFORMATION: /product= "UL147"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (1368..1721)
; OTHER INFORMATION: /product= "UL152"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (2504..3337)
; OTHER INFORMATION: /product= "UL153"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (3515..4711)
; OTHER INFORMATION: /product= "UL154"
; US-09-253-682-1
Alignment Scores:
Pred. No.: 235 Length: 4711
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 3 Gaps: 0
US-09-854-133-587 (1-16) x US-09-253-682-1 (1-4711)
QY 9 PheilleillePheTrpIle 14
|||||
```

Db 4654 TTTATTATTTTGGATT 4637

RESULT 26

US-09-527-657-1/c

; Sequence 1, Application US/09527657

; Patent No. 6291236

; GENERAL INFORMATION:

; APPLICANT: Spaete, Richard

; Cha, Tai-An

; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Luann Cserri Attorney at Law

; STREET: 750 Arimo Avenue

; CITY: Oakland

; STATE: CA

; COUNTRY: USA

; ZIP: 94610

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/527,657

; FILING DATE: 17-Mar-2000

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/926,922

; FILING DATE: September 10, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Cserri, Luann

; REGISTRATION NUMBER: 31,822

; REFERENCE/DOCKET NUMBER: AVIR 11A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 510-834-1448

; TELEFAX: 510-839-7810

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4711 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Human CMV

; STRAIN: Towne

; FEATURE:

; NAME/KEY: CDS

; LOCATION: complement (845..1321)

; OTHER INFORMATION: /product= "UL147"

; FEATURE:

; NAME/KEY: CDS

; LOCATION: complement (1368..1721)

; OTHER INFORMATION: /product= "UL152"

; FEATURE:

; NAME/KEY: CDS

; LOCATION: complement (2504..3337)

; OTHER INFORMATION: /product= "UL153"

; FEATURE:

; NAME/KEY: CDS

; LOCATION: complement (3515..4711)

; OTHER INFORMATION: /product= "UL154"

; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-527-657-1

Alignment Scores:

Pred. No.:	235	Length:	4711
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

Query Match: 37.50% Indels: 0

DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-527-657-1 (1-4711)

QY 9 PheillellepHeTrpIle 14

|||||

Db 4654 TTTATTATTTTGGATT 4637

RESULT 27

US-08-936-107A-8/c

; Sequence 8, Application US/08936107A

; Patent No. 6403306

; GENERAL INFORMATION:

; APPLICANT: Stephens, David S.

; APPLICANT: Swartley, John S.

; TITLE OF INVENTION: Serogroup-Specific Nucleotide Sequences

; TITLE OF INVENTION: in the Molecular Typing of Bacterial Isolates and the

; TITLE OF INVENTION: Preparation of Vaccines Thereto

; NUMBER OF SEQUENCES: 51

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

; STREET: 5370 Manhattan Circle, Suite 201

; CITY: Boulder

; STATE: Colorado

; COUNTRY: US

; ZIP: 80303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/936,107A

; FILING DATE: 23-SEP-1997

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/827,622

; FILING DATE: 09-APR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Caruthers, Jennie M.

; REGISTRATION NUMBER: 34,464

; REFERENCE/DOCKET NUMBER: 77-97

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 499-8080

; TELEFAX: (303) 499-8089

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5064 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: not relevant

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 479..1597

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1599..3236

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 3309..4052

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 4054..4917

US-08-936-107A-8

Alignment Scores:

Pred. No.:	250	Length:	5064
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

Query Match: 37.50% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-08-936-107A-8 (1-5064)

QY 10 IleilepHeTrpillephe 15  
|||||

Db 3530 ATCATCTTCTGGATATTT 3513

## RESULT 28

US-08-961-527-98/c

; Sequence 98, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,527

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 98:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 8654 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; US-08-961-527-98

Alignment Scores:

Pred. No.: 399

Score: 6.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 37.50%

DB: 4

US-09-854-133-587 (1-16) x US-08-961-527-98 (1-8654)

QY 5 CysGlyIleAspPheille 10

|||||

Db 233 TCGGGCATTGATTCATC 216

## RESULT 29

US-08-961-527-191

; Sequence 191, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,527

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 191:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 10357 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; US-08-961-527-191

Alignment Scores:

Pred. No.: 466

Score: 6.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 37.50%

DB: 4

US-09-854-133-587 (1-16) x US-08-961-527-191 (1-10357)

QY 7 IleAspPheillellephe 12

|||||

Db 9885 ATTGATTTTATATCTTT 9902

## RESULT 30

US-08-370-975B-6/c

; Sequence 6, Application US/08370975B

; Patent No. 5622851

; GENERAL INFORMATION:

; APPLICANT: Maley, Frank

; APPLICANT: Maley, Gladys F.

; APPLICANT: Weiner, Karen X.B.

; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nixon, Hargrave, Devans & Doyle

; STREET: Clinton Square, P.O. Box 1051

; CITY: Rochester

; STATE: New York

; COUNTRY: USA

; ZIP: 14603

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/370,975B

; FILING DATE: 10-JAN-1995

Length: 10357

Matches: 6

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20894/80
; TELEPHONE: (716)263-1636
; TELEFAX: (716)263-1600
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 4q35
US-08-370-975B-6

Alignment Scores:
Pred. No.: 836 Length: 20303
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 1 Gaps: 0

US-09-854-133-587 (1-16) x US-08-370-975B-6 (1-20303)

QY 10 IleIlePheTrpIlePhe 15
Db 4242 ATTATCTTTGGATT TTT 4225

RESULT 31
US-08-370-975B-1/c
; Sequence 1, Application US/08370975B
; Patent No. 5622851
; GENERAL INFORMATION:
; APPLICANT: Maley, Frank
; APPLICANT: Maley, Gladys F.
; APPLICANT: Weiner, Karen X.B.
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,975B
; FILING DATE: 10-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20894/80
; TELEPHONE: (716)263-1636
; TELEFAX: (716)263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26764 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

```
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 4q35
US-08-370-975B-1

Alignment Scores:
Pred. No.: 1.06e+03 Length: 26764
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 1 Gaps: 0

US-09-854-133-587 (1-16) x US-08-370-975B-1 (1-26764)

QY 10 IleIlePheTrpIlePhe 15
Db 6205 ATTATCTTTGGATT TTT 6188

RESULT 32
US-09-341-444A-28
; Sequence 28, Application US/09341444A
; Patent No. 6440666
; GENERAL INFORMATION:
; APPLICANT: Groenen, Martinus Antonius Mathilda
; APPLICANT: Albers, Gerardus Antonius Arnoldus
; TITLE OF INVENTION: Selection For Dwarfism in Poultry
; FILE REFERENCE: 310-1009
; CURRENT APPLICATION NUMBER: US/09/341,444A
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/NL98/00021
; PRIOR FILING DATE: 1998-01-12
; PRIOR APPLICATION NUMBER: EP 97200070.7
; PRIOR FILING DATE: 1997-01-10
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28
; LENGTH: 24
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: OTHER INFORMATION:Oligonucleotide primer
US-09-341-444A-28

Alignment Scores:
Pred. No.: 33.4 Length: 24
Score: 5.00 Matches: 5
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.25% Indels: 0
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-341-444A-28 (1-24)

QY 7 IleAspPheIlelle 11
Db 6 ATTGATTTCATCATT 20

RESULT 33
US-08-985-659-2/c
; Sequence 2, Application US/08985659
; Patent No. 6211134
; GENERAL INFORMATION:
; APPLICANT: Caldwell, Robert M
; APPLICANT: Mitchinson, Colin
; APPLICANT: Ropp, Traci M
; TITLE OF INVENTION: Mutant Alpha-Amylase
; FILE REFERENCE: A-69396/DJB/DAV/JJD
; CURRENT APPLICATION NUMBER: US/08/985,659
; CURRENT FILING DATE: 1997-12-09
; PRIOR APPLICATION NUMBER: 08/645,971
; PRIOR FILING DATE: 1996-05-14
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
```



; SEQ ID NO 2  
; LENGTH: 48  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-08-985-659-2  
  
Alignment Scores: 61 Length: 48  
Pred. No.: 5.00 Matches: 5  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 31.25% Gaps: 0  
DB: 4  
  
US-09-854-133-587 (1-16) x US-08-985-659-2 (1-48)  
  
QY 9 PheileilePheTrp 13  
Db 45 TTTATTATTTTGG 31  
  
RESULT 34  
US-09-046-247-138  
; Sequence 138, Application US/09046247  
; Patent No. 6124449  
; GENERAL INFORMATION:  
; APPLICANT: NIKOS PAGRATIS  
; APPLICANT: LARRY GOLD  
; TITLE OF INVENTION: HIGH AFFINITY TGF? NUCLEIC  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Swanson and Bratschun, L.L.C.  
; STREET: 8400 East Prentice Avenue, Suite #200  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS DOS  
; SOFTWARE: Word 7.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/046,247  
; FILING DATE: 23-MARCH-1998  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/458,424  
; FILING DATE: 2-JUNE-1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/714,131  
; FILING DATE: 10-JUNE-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/536,428  
; FILING DATE: 11-JUNE-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/964,624  
; FILING DATE: 21-OCTOBER-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/117,991  
; FILING DATE: 8-SEPTEMBER-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/931,473  
; FILING DATE: 17-AUGUST-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Barry Swanson  
; REGISTRATION NUMBER: 33,215  
; REFERENCE/DOCKET NUMBER: NEX 34.2/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 793-3333  
; TELEFAX: (303) 793-3433

; INFORMATION FOR SEQ ID NO: 138:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 66 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: RNA  
; FEATURE:  
; OTHER INFORMATION: All pyrimidines are 2'-F modified  
US-09-046-247-138  
  
Alignment Scores: 80.4 Length: 66  
Pred. No.: 5.00 Matches: 5  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 31.25% Gaps: 0  
DB: 3  
  
US-09-854-133-587 (1-16) x US-09-046-247-138 (1-66)  
  
QY 5 CysGlyIleAspPhe 9  
Db 11 UGCGGUUUGACUUU 25  
  
RESULT 35  
US-08-450-834-4  
; Sequence 4, Application US/08450834  
; Patent No. 5773705  
; GENERAL INFORMATION:  
; APPLICANT: Vierstra, Richard D  
; APPLICANT: Hondred, David  
; APPLICANT: Callis, Judy  
; TITLE OF INVENTION: Ubiquitin Fusion Protein System for  
; TITLE OF INVENTION: Protein Production in Plants  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: P.O. Box 2113  
; CITY: Madison  
; STATE: WI  
; COUNTRY: USA  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/450,834  
; FILING DATE: 25-MAY-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/999,709  
; FILING DATE: 31-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 960296.92425  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: UBQ-GUS

```
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4..9
; OTHER INFORMATION: /function= "Sac II restriction
; OTHER INFORMATION: site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 39..44
; OTHER INFORMATION: /function= "Xho I restriction site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 103..108
; OTHER INFORMATION: /function= "Bcl I restriction site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 109..114
; OTHER INFORMATION: /function= "Xba I restriction site"
;
US-08-450-834-4
```

```
Alignment Scores:
Pred. No.: 132 Length: 117
Score: 5.00 Matches: 5
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.25% Indels: 0
DB: 1 Gaps: 0
```

US-09-854-133-587 (1-16) x US-08-450-834-4 (1-117)

```
Qy 4 AsnCysGlyIleAsp 8
      |||||
Db 92 AACTGTGGATTGAT 106
```

## RESULT 36

```
US-09-134-001C-548/c
; Sequence 548, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
```

```
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 548
; LENGTH: 222
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-548
```

```
Alignment Scores:
Pred. No.: 231 Length: 222
Score: 5.00 Matches: 5
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.25% Indels: 0
DB: 4 Gaps: 0
```

US-09-854-133-587 (1-16) x US-09-134-001C-548 (1-222)

```
Qy 11 IlePheTrpIlePhe 15
      |||||
Db 164 ATTTCTGGATCTTC 150
```

## RESULT 37

```
US-09-134-001C-2829
; Sequence 2829, Application US/09134001C
; Patent No. 6380370
```

```
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2829
; LENGTH: 225
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2829
```

```
Alignment Scores:
Pred. No.: 233 Length: 225
Score: 5.00 Matches: 5
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.25% Indels: 0
DB: 4 Gaps: 0
```

US-09-854-133-587 (1-16) x US-09-134-001C-2829 (1-225)

```
Qy 9 PheIleIlePheTrp 13
      |||||
Db 118 TTTATAATTTTTGG 132
```

## RESULT 38

```
US-08-410-804-6
; Sequence 6, Application US/08410804
; Patent No. 5632994
; GENERAL INFORMATION:
```

```
; APPLICANT: Reed, John C.
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cathryn Campbell
; STREET: 4370 La Jolla Village Drive, Ste 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,804
; FILING DATE: 27-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/259,514
; FILING DATE: 14-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1389
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
```

```
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

```

; MOLECULE TYPE: CDNA
US-08-410-804-6

Alignment Scores: 263 Length: 258
Pred. No.: 5.00 Matches: 5
Score: 100.00% Conservatives: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 31.25% Gaps: 0
DB: 1

US-09-854-133-587 (1-16) x US-08-410-804-6 (1-258)

QY 7 IleAspPheIleIle 11
Db 172 ATAGATTTCATAATT 186

RESULT 39
US-08-259-514-6
; Sequence 6, Application US/08259514
; Patent No. 5747245
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cathryn Campbell
; STREET: 4370 La Jolla Village Drive. Ste 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/259,514
; FILING DATE: 14-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/410,804
; FILING DATE: 27-MAR-1995
; APPLICATION NUMBER: US 08/259,514
; FILING DATE: 14-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9954
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-259-514-6

Alignment Scores: 263 Length: 258
Pred. No.: 5.00 Matches: 5
Score: 100.00% Conservatives: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 31.25% Gaps: 0
DB: 1

US-09-854-133-587 (1-16) x US-08-259-514-6 (1-258)

QY 7 IleAspPheIleIle 11
Db 172 ATAGATTTCATAATT 186

RESULT 40
US-08-259-514-6
; Sequence 6, Application US/08259514
; Patent No. 5747245
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cathryn Campbell
; STREET: 4370 La Jolla Village Drive. Ste 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/259,514
; FILING DATE: 14-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/410,804
; FILING DATE: 27-MAR-1995
; APPLICATION NUMBER: US 08/259,514
; FILING DATE: 14-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9954
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-259-514-6

Alignment Scores: 263 Length: 258
Pred. No.: 5.00 Matches: 5
Score: 100.00% Conservatives: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 31.25% Gaps: 0
DB: 1

US-09-854-133-587 (1-16) x US-08-259-514-6 (1-258)

QY 7 IleAspPheIleIle 11
Db 172 ATAGATTTCATAATT 186

RESULT 41
US-09-438-906-15/c
; Sequence 15, Application US/09438906
; Patent No. 6465185
; GENERAL INFORMATION:
; APPLICANT: Goldfine, Ira
; APPLICANT: Trischitta, Vincenzo
; APPLICANT: Pizzuti, Antonio
; APPLICANT: Vigneri, Riccardo
; APPLICANT: Frittitta, Lucia
; TITLE OF INVENTION: Polymorphic Human PC-1 Sequences
; FILE REFERENCE: 9076-089
; CURRENT APPLICATION NUMBER: US/09/438,906
; CURRENT FILING DATE: 1999-11-18

```

; PRIOR APPLICATION NUMBER: 60/108,853  
; PRIOR FILING DATE: 1998-11-18  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 313  
; TYPE: DNA  
; ORGANISM: H. sapiens  
US-09-438-906-15

Alignment Scores:  
Pred. No.: 311 Length: 313  
Score: 5.00 Matches: 5  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 31.25% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-438-906-15 (1-313)

QY 8 AspPheillellePhe 12  
| | | | | | | | | |  
Db 215 GATTTTATTATTTT 201

## RESULT 42

US-09-134-001C-2354/C  
; Sequence 2354, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2354  
; LENGTH: 390  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2354

## Alignment Scores:

Pred. No.: 376 Length: 390  
Score: 5.00 Matches: 5  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 31.25% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-134-001C-2354 (1-390)

QY 7 IleAspPheillelle 11  
| | | | | | | | | |  
Db 146 ATTGACTTCATCATA 132

## RESULT 43

US-09-134-001C-2395  
; Sequence 2395, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2395  
; LENGTH: 405  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2395

Alignment Scores:  
Pred. No.: 389 Length: 405  
Score: 5.00 Matches: 5  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 31.25% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-134-001C-2395 (1-405)

QY 7 IleAspPheillelle 11  
| | | | | | | | | |  
Db 265 ATTGACTTCATCATA 279

## RESULT 44

US-08-469-667-21/C  
; Sequence 21, Application US/08469667  
; Patent No. 5733748  
; GENERAL INFORMATION:

; APPLICANT: Yu, Guo-Liang  
; APPLICANT: Rosen, Craig  
; TITLE OF INVENTION: Colon Specific Genes and Proteins  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,  
; ADDRESSEE: Stewart & Olstein  
; STREET: 6 Becker Farm Road  
; CITY: Roseland  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07068-1739  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/469,667  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferraro, Gregory D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-435  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 409 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-469-667-21

## Alignment Scores:

Pred. No.: 392 Length: 409  
Score: 5.00 Matches: 5  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 31.25% Indels: 0  
DB: 1 Gaps: 0



us-09-854-133-587-133-587.oligo.rni

Tue May 13 12:12:58 2003

US-09-854-133-587 (1-16) x US-08-469-667-21 (1-409)

QY 11 IlePheTrpIlePhe 15  
 Db 171 ATTTCTGGATATT 157

RESULT 45

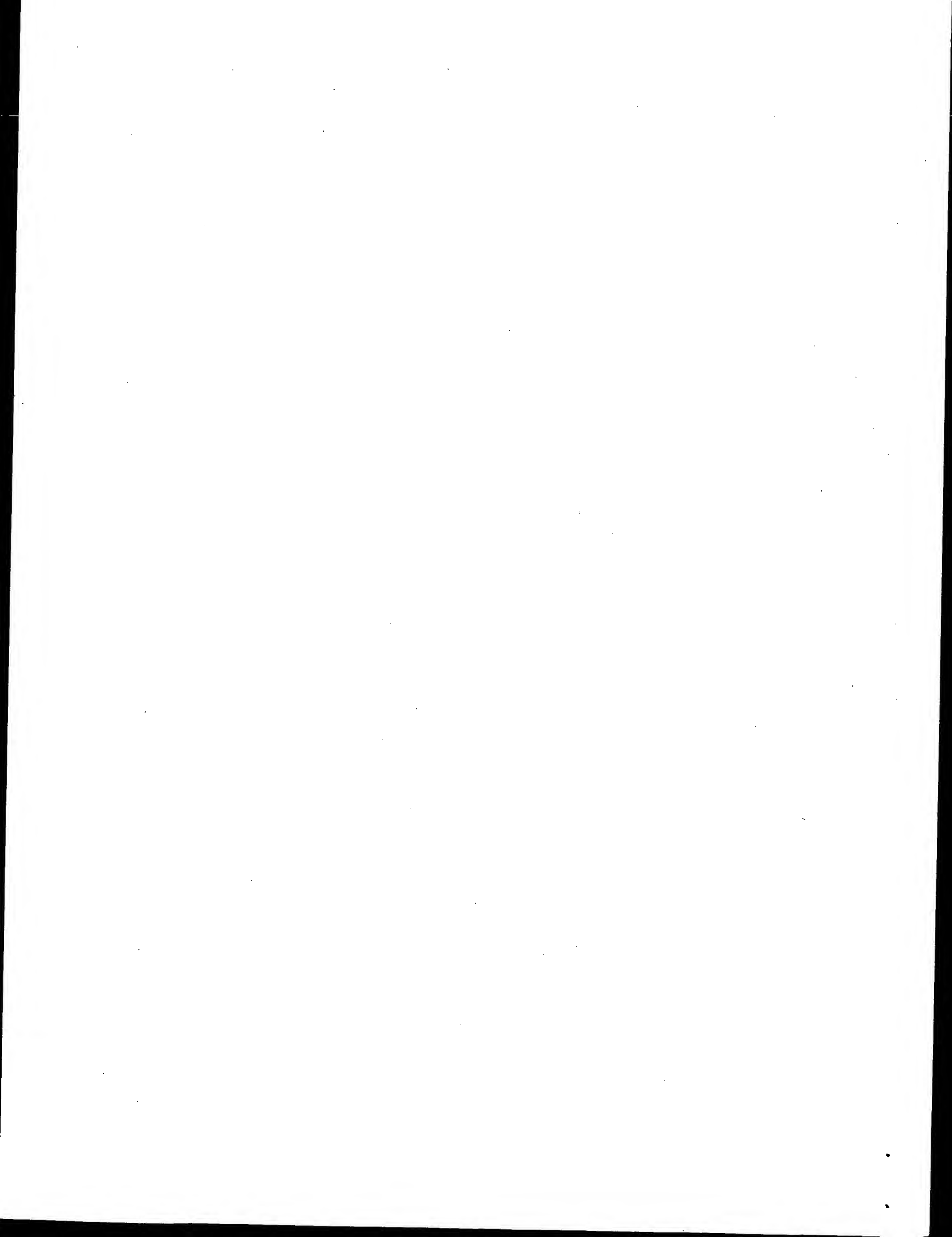
US-09-224-110-21/c  
 ; Sequence 21, Application US/09224110  
 ; Patent No. 6337195  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yu, Guo-Liang  
 ; APPLICANT: Rosen, Craig  
 ; TITLE OF INVENTION: Colon Specific Genes and Proteins  
 ; NUMBER OF SEQUENCES: 24  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,  
 ; ADDRESSEE: Stewart & Olstein  
 ; STREET: 6 Becker Farm Road  
 ; CITY: Roseland  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 07068-1739  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/224,110  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/469,667  
 ; FILING DATE: 06-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ferraro, Gregory D.  
 ; REGISTRATION NUMBER: 36,134  
 ; REFERENCE/DOCKET NUMBER: 325800-435  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 201-994-1700  
 ; TELEFAX: 201-994-1744  
 ; INFORMATION FOR SEQ ID NO: 21:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 409 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; US-09-224-110-21

Alignment Scores:  
 Pred. No.: 392 Length: 409  
 Score: 5.00 Matches: 5  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 31.25% Indels: 0  
 DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-224-110-21 (1-409)

QY 11 IlePheTrpIlePhe 15  
 Db 171 ATTTCTGGATATT 157

Search completed: May 11, 2003, 17:17:52  
 Job time : 18.9204 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 17:18:00 ; Search time 18.9735 Seconds  
(without alignments)  
1047.953 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FQANGCIDFIIFWIFW 16

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 783854 seqs, 621352466 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 1563015

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 60 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_174135\_809/app\_query.fasta\_1.462  
-DB=Published\_Applications\_NA -QFMT=fastap -SUFFIX=oligo.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=60 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1  
-ALIGN=45 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09854133@cgn1\_1\_117@runat\_05052003\_174135\_809  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published\_Applications\_NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	100.0	337	9 US-09-854-133-442	Sequence 442, App
2	16	100.0	337	10 US-09-738-973-442	Sequence 442, App
3	16	100.0	2239	9 US-09-854-133-440	Sequence 440, App
4	16	100.0	2239	10 US-09-738-973-440	Sequence 440, App

ALIGNMENTS

RESULT 1

US-09-854-133-442  
; Sequence 442, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:

; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10

5	16	100.0	5981	9	US-09-854-133-441	Sequence 441, App
6	16	100.0	5981	10	US-09-738-973-441	Sequence 441, App
c 7	7	43.8	170834	10	US-09-835-232-7	Sequence 7, Appli
8	6	37.5	26	10	US-09-006-298-7	Sequence 7, Appli
c 9	6	37.5	250	10	US-09-867-701-51	Sequence 51, Appl
10	6	37.5	277	9	US-10-076-622-101	Sequence 101, App
11	6	37.5	277	10	US-09-604-287A-101	Sequence 101, App
12	6	37.5	277	10	US-09-339-338-101	Sequence 101, App
13	6	37.5	277	12	US-10-007-805-101	Sequence 101, App
14	6	37.5	283	10	US-09-867-701-10058	Sequence 10058, A
15	6	37.5	390	9	US-09-796-692-9005	Sequence 9005, Ap
16	6	37.5	390	9	US-10-040-862-9005	Sequence 9005, Ap
c 17	6	37.5	416	9	US-09-918-995-5321	Sequence 5321, Ap
c 18	6	37.5	462	9	US-09-918-995-13913	Sequence 13913, A
19	6	37.5	498	10	US-09-880-107-1080	Sequence 1080, Ap
20	6	37.5	505	10	US-09-998-598-1101	Sequence 1101, Ap
21	6	37.5	579	10	US-09-764-877-243	Sequence 243, App
22	6	37.5	600	10	US-09-864-761-8345	Sequence 8345, Ap
23	6	37.5	696	10	US-09-974-300-6399	Sequence 6399, Ap
c 24	6	37.5	989	10	US-09-867-550-1741	Sequence 1741, Ap
25	6	37.5	1188	10	US-09-741-669-167	Sequence 167, App
26	6	37.5	1260	10	US-09-815-242-4288	Sequence 4288, Ap
c 27	6	37.5	1263	10	US-09-815-242-8550	Sequence 8550, Ap
c 28	6	37.5	1263	10	US-09-815-242-8719	Sequence 8719, Ap
29	6	37.5	1268	10	US-09-070-927A-557	Sequence 557, App
30	6	37.5	1330	9	US-09-746-660A-11	Sequence 11, Appl
31	6	37.5	1455	9	US-10-142-231-43	Sequence 43, Appl
32	6	37.5	1455	9	US-10-142-231-44	Sequence 44, Appl
33	6	37.5	1491	9	US-09-746-660A-9	Sequence 9, Appli
34	6	37.5	1524	9	US-10-142-231-52	Sequence 52, Appl
c 35	6	37.5	1623	10	US-09-974-300-6367	Sequence 6367, Ap
c 36	6	37.5	1786	10	US-09-764-877-4007	Sequence 4007, Ap
c 37	6	37.5	2000	9	US-09-938-842A-3762	Sequence 3762, Ap
38	6	37.5	2000	9	US-09-938-842A-3983	Sequence 3983, Ap
39	6	37.5	2000	9	US-09-938-842A-4984	Sequence 4984, Ap
c 40	6	37.5	2524	7	US-08-781-986A-360	Sequence 360, App
c 41	6	37.5	2692	9	US-10-092-154-1254	Sequence 1254, Ap
c 42	6	37.5	2692	10	US-09-764-847-1254	Sequence 1254, Ap
c 43	6	37.5	2694	10	US-10-092-154-1255	Sequence 1255, Ap
c 44	6	37.5	2694	10	US-09-764-847-1255	Sequence 1255, Ap
45	6	37.5	2912	9	US-10-071-766-61	Sequence 61, Appl
46	6	37.5	3720	9	US-10-076-622-576	Sequence 576, App
47	6	37.5	3720	12	US-10-007-805-576	Sequence 576, App
48	6	37.5	5285	9	US-10-239-676-157	Sequence 157, App
49	6	37.5	5508	10	US-09-954-456-943	Sequence 943, App
c 50	6	37.5	11230	9	US-10-150-821-3	Sequence 3, Appli
c 51	6	37.5	11230	10	US-09-911-842-3	Sequence 3, Appli
c 52	6	37.5	22786	10	US-09-764-877-3727	Sequence 3727, Ap
53	6	37.5	27332	9	US-10-092-154-1555	Sequence 1555, Ap
54	6	37.5	27332	10	US-09-764-847-1555	Sequence 1555, Ap
55	6	37.5	55795	10	US-09-880-107-1543	Sequence 1543, Ap
56	6	37.5	124884	9	US-09-913-514-1	Sequence 1, Appli
57	6	37.5	125157	9	US-09-913-514-2	Sequence 2, Appli
58	6	37.5	640681	10	US-09-790-988-1	Sequence 1, Appli
c 59	5	31.2	17	9	US-09-818-875-1906	Sequence 1906, Ap
60	5	31.2	17	9	US-09-818-875-1907	Sequence 1907, Ap

```
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 442
; LENGTH: 337
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-133-442
```

```
Alignment Scores:
Pred. No.: 1.11e-09 Length: 337
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
```

US-09-854-133-587 (1-16) x US-09-854-133-442 (1-337)

```
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
Db 107 TTCCAGGCCCAATTGTGGCATAGATTTTATCATATCTGGATTTTGG 154
```

## RESULT 2

```
US-09-738-973-442
; Sequence 442, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
```

```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 442
; LENGTH: 337
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-738-973-442
```

```
Alignment Scores:
Pred. No.: 1.11e-09 Length: 337
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0
```

US-09-854-133-587 (1-16) x US-09-738-973-442 (1-337)

```
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
Db 107 TTCCAGGCCCAATTGTGGCATAGATTTTATCATATCTGGATTTTGG 154
```

## RESULT 3

```
US-09-854-133-440
; Sequence 440, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 440
; LENGTH: 2239
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-133-440
```

## Alignment Scores:

```
Pred. No.: 7.04e-09 Length: 2239
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
```

US-09-854-133-587 (1-16) x US-09-854-133-440 (1-2239)

```
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
Db 104 TTCCAGGCCCAATTGTGGCATAGATTTTATCATATCTGGATTTTGG 151
```

## RESULT 4

```
US-09-738-973-440
; Sequence 440, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
```

```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
```

```
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 440
; LENGTH: 2239
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-738-973-440
```

## Alignment Scores:

```
Pred. No.: 7.04e-09 Length: 2239
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0
```

US-09-854-133-587 (1-16) x US-09-738-973-440 (1-2239)

```
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
```



|||||  
Db 104 TTCCAGGCCAATTGTGGCATAGATTTTATCATATCTGGATTTTGG 151

## RESULT 5

US-09-854-133-441  
; Sequence 441, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 441  
; LENGTH: 5981  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-854-133-441

Alignment Scores: Length: 5981  
Pred. No.: 1.84e-08 Matches: 16  
Score: 16.00 Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 100.00% Gaps: 0  
DB:

US-09-854-133-587 (1-16) x US-09-854-133-441 (1-5981)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
|||||  
Db 102 TTCCAGGCCAATTGTGGCATAGATTTTATCATATCTGGATTTTGG 149

## RESULT 6

US-09-738-973-441  
; Sequence 441, Application US/09738973  
; Patent No. US20020110563A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Indirias, Carol Yoseph  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Elliot, Mark  
; APPLICANT: Mannion, Jane  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C9  
; CURRENT APPLICATION NUMBER: US/09/738,973  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 587  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 441  
; LENGTH: 5981  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-738-973-441

Alignment Scores: Length: 5981  
Pred. No.: 1.84e-08 Matches: 16  
Score: 16.00

Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-738-973-441 (1-5981)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
|||||  
Db 102 TTCCAGGCCAATTGTGGCATAGATTTTATCATATCTGGATTTTGG 149

## RESULT 7

US-09-835-232-7/c  
; Sequence 7, Application US/09835232  
; Patent No. US20020098489A1  
; GENERAL INFORMATION:  
; APPLICANT: Leder, Philip  
; APPLICANT: Leader, Benjamin  
; TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: 00383/052002  
; CURRENT APPLICATION NUMBER: US/09/835,232  
; CURRENT FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: US 60/196,811  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 170834  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(170834)  
; OTHER INFORMATION: n= A,T,C, or G  
US-09-835-232-7

Alignment Scores: Length: 170834  
Pred. No.: 3.32e+03 Matches: 7  
Score: 7.00 Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 43.75% Gaps: 0  
DB:

US-09-854-133-587 (1-16) x US-09-835-232-7 (1-170834)

QY 10 IleIlePheTrpIlePheTrp 16  
|||||  
Db 95470 ATCATCTTTTGGATTTCTGG 95450

## RESULT 8

US-09-006-298-7  
; Sequence 7, Application US/09006298  
; Patent No. US20020082224A1  
; GENERAL INFORMATION:  
; APPLICANT: Jolly, Douglas J.  
; APPLICANT: Moore, Margaret D.  
; APPLICANT: Chada, Sunil  
; TITLE OF INVENTION: NON-IMMUNOGENIC PRODRUGS AND SELECTABLE  
; TITLE OF INVENTION: MARKERS FOR USE IN GENE THERAPY  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

```
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/006,298
/ FILING DATE: 13-JAN-1998
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: McMasters, David D.
/ REGISTRATION NUMBER: 33,963
/ REFERENCE/DOCKET NUMBER: 930049.459
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 26 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
US-09-006-298-7
```

```
Alignment Scores:
Pred. No.: 7.68 Length: 26
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0
```

US-09-854-133-587 (1-16) x US-09-006-298-7 (1-26)

```
QY 6 GlyIleAspPheIleIle 11
Db 4 GGGATCGATTTCATCATC 21
```

## RESULT 9

```
US-09-867-701-51/c
/ Sequence 51, Application US/09867701
/ Patent No. US20020132237A1
/ GENERAL INFORMATION:
/ APPLICANT: Aglate, Paul A.
/ APPLICANT: Jones, Robert
/ APPLICANT: Harlocker, Susan L.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
/ FILE REFERENCE: 210121.497
/ CURRENT APPLICATION NUMBER: US/09/867,701
/ CURRENT FILING DATE: 2001-05-29
/ NUMBER OF SEQ ID NOS: 10912
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 51
/ LENGTH: 250
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-867-701-51
```

```
Alignment Scores:
Pred. No.: 70.1 Length: 250
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0
```

US-09-854-133-587 (1-16) x US-09-867-701-51 (1-250)

```
QY 10 IleIlePheTrpIlePhe 15
Db 100 ATAATTTCCTGGATTTT 83
```

## RESULT 10

```
US-10-076-622-101
/ Sequence 101, Application US/10076622
/ Publication No. US20030023036A1
```

```
/ GENERAL INFORMATION:
/ APPLICANT: Houghton, Raymond L.
/ APPLICANT: Sleath, Paul R.
/ APPLICANT: Persing, David H.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
/ FILE REFERENCE: 210121.470C11
/ CURRENT APPLICATION NUMBER: US/10/076,622
/ CURRENT FILING DATE: 2002-02-13
/ NUMBER OF SEQ ID NOS: 627
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 101
/ LENGTH: 277
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-076-622-101
```

```
Alignment Scores:
Pred. No.: 77.5 Length: 277
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 9 Gaps: 0
```

US-09-854-133-587 (1-16) x US-10-076-622-101 (1-277)

```
QY 10 IleIlePheTrpIlePhe 15
Db 96 ATTATATTTGGATCTTC 113
```

## RESULT 11

```
US-09-604-287A-101
/ Sequence 101, Application US/09604287A
/ Patent No. US20020064872A1
/ GENERAL INFORMATION:
/ APPLICANT: Jiang, Yugu
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Hepler, William T.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.470C7
/ CURRENT APPLICATION NUMBER: US/09/604,287A
/ CURRENT FILING DATE: 2000-06-22
/ NUMBER OF SEQ ID NOS: 489
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 101
/ LENGTH: 277
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-604-287A-101
```

```
Alignment Scores:
Pred. No.: 77.5 Length: 277
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0
```

US-09-854-133-587 (1-16) x US-09-604-287A-101 (1-277)

```
QY 10 IleIlePheTrpIlePhe 15
Db 96 ATTATATTTGGATCTTC 113
```

## RESULT 12

```
US-09-339-338-101
/ Sequence 101, Application US/09339338A
/ Patent No. US20020102602A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Yuqiu, Jiang
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, JIANGchun
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND METHODS FOR
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR
; FILE REFERENCE: 210121.470C2
; CURRENT APPLICATION NUMBER: US/09/339,338A
; CURRENT FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 315
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 101
; LENGTH: 277
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-339-338-101

Alignment Scores:
Pred. No.: 77.5 Length: 277
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-339-338-101 (1-277)

QY 10 IleIlePheTrpIlePhe 15
      |||||||
Db 96 ATTATATTTGGATCTTC 113

RESULT 13
US-10-007-805-101
; Sequence 101, Application US/10007805
; Patent No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, JIANGchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 277
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-007-805-101

Alignment Scores:
Pred. No.: 77.5 Length: 277
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x US-10-007-805-101 (1-277)

QY 10 IleIlePheTrpIlePhe 15

```

```
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9005
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-9005
```

```
Alignment Scores:
Pred. No.: 108
Score: 6.00
Length: 390
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 37.50%
DB: 9
Matches: 6
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0
```

```
US-09-854-133-587 (1-16) x US-09-796-692-9005 (1-390)
```

```
QY 10 IleIlePheTrpIlePhe 15
Db 229 ATTATATTGGATCTTC 246
|||||
```

## RESULT 16

```
US-10-040-862-9005
; Sequence 9005, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
```

```
; APPLICANT: Corixa Corporation
```

```
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
```

```
; FILE REFERENCE: 014058-013520US
```

```
; CURRENT APPLICATION NUMBER: US/10/040,862
```

```
; CURRENT FILING DATE: 2001-11-06
```

```
; PRIOR APPLICATION NUMBER: US 60/186,126
```

```
; PRIOR FILING DATE: 2000-03-01
```

```
; PRIOR APPLICATION NUMBER: US 60/190,479
```

```
; PRIOR FILING DATE: 2000-03-17
```

```
; PRIOR APPLICATION NUMBER: US 60/200,545
```

```
; PRIOR FILING DATE: 2000-04-27
```

```
; PRIOR APPLICATION NUMBER: US 60/200,303
```

```
; PRIOR FILING DATE: 2000-04-28
```

```
; PRIOR APPLICATION NUMBER: US 60/200,779
```

```
; PRIOR FILING DATE: 2000-04-28
```

```
; PRIOR APPLICATION NUMBER: US 60/200,999
```

```
; PRIOR FILING DATE: 2000-05-01
```

```
; PRIOR APPLICATION NUMBER: US 60/202,084
```

```
; PRIOR FILING DATE: 2000-05-04
```

```
; PRIOR APPLICATION NUMBER: US 60/206,201
```

```
; PRIOR FILING DATE: 2000-05-22
```

```
; PRIOR APPLICATION NUMBER: US 60/218,950
```

```
; PRIOR FILING DATE: 2000-07-14
```

```
; PRIOR APPLICATION NUMBER: US 60/222,903
```

```
; PRIOR FILING DATE: 2000-08-03
```

```
; PRIOR APPLICATION NUMBER: US 60/223,416
```

```
; PRIOR FILING DATE: 2000-08-04
```

```
; PRIOR APPLICATION NUMBER: US 60/223,378
```

```
; PRIOR FILING DATE: 2000-08-07
```

```
; PRIOR APPLICATION NUMBER: US 09/796,692
```

```
; PRIOR FILING DATE: 2001-03-01
```

```
; NUMBER OF SEQ ID NOS: 10467
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 9005
```

```
; LENGTH: 390
```

```
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
```

```
US-10-040-862-9005
```

## Alignment Scores:

```
Pred. No.: 108
Score: 6.00
Length: 390
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 37.50%
DB: 9
Matches: 6
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0
```

```
US-09-854-133-587 (1-16) x US-10-040-862-9005 (1-390)
```

```
QY 10 IleIlePheTrpIlePhe 15
Db 229 ATTATATTGGATCTTC 246
|||||
```

## RESULT 17

```
US-09-918-995-5321/c
```

```
; Sequence 5321, Application US/09918995
```

```
; Publication No. US20030073623A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Hyseq, Inc.
```

```
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
```

```
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
```

```
; FILE REFERENCE: 20411-756
```

```
; CURRENT APPLICATION NUMBER: US/09/918,995
```

```
; CURRENT FILING DATE: 2001-07-30
```

```
; PRIOR APPLICATION NUMBER: US/09/235,076
```

```
; PRIOR FILING DATE: 1999-01-20
```

```
; NUMBER OF SEQ ID NOS: 38054
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 5321
```

```
; LENGTH: 416
```

```
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
```

```
; FEATURE:
```

```
; NAME/KEY: misc_feature
```

```
; LOCATION: (1)...(416)
```

```
; OTHER INFORMATION: n = A,T,C or G
```

```
US-09-918-995-5321
```

## Alignment Scores:

```
Pred. No.: 115
Score: 6.00
Length: 416
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 37.50%
DB: 9
Matches: 6
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0
```

```
US-09-854-133-587 (1-16) x US-09-918-995-5321 (1-416)
```

```
QY 6 GlyIleAspPheIle 11
Db 105 GGCATTGACTTCATTATA 88
|||||
```

## RESULT 18

```
US-09-918-995-13913/c
```

```
; Sequence 13913, Application US/09918995
```

```
; Publication No. US20030073623A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Hyseq, Inc.
```

```
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
```

```
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
```

```
; FILE REFERENCE: 20411-756
```

```
; CURRENT APPLICATION NUMBER: US/09/918,995
```

```
; CURRENT FILING DATE: 2001-07-30
```

```
; PRIOR APPLICATION NUMBER: US/09/235,076
```

```
; PRIOR FILING DATE: 1999-01-20
```

```
; NUMBER OF SEQ ID NOS: 38054
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 13913
```

```
; LENGTH: 462
```

```
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
```

```
; FEATURE:
```

```
; NAME/KEY: misc_feature
```



```
; LOCATION: (1)...(462)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-13913

Alignment Scores:
Pred. No.: 128 Length: 462
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-918-995-13913 (1-462)

QY 10 IleilepHeTrpIlePhe 15
| | | | | | | | | | | | | | | |
Db 351 ATTATTTCTGGATATTC 334

RESULT 19
US-09-880-107-1080
; Sequence 1080, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1080
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA443756
US-09-880-107-1080

Alignment Scores:
Pred. No.: 137 Length: 498
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-880-107-1080 (1-498)

QY 6 GlyIleAspPheIleIle 11
| | | | | | | | | | | | | | | |
Db 48 GGCATTGACTTCATTATA 65

RESULT 20
US-09-998-598-1101
; Sequence 1101, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16

; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 1101
; LENGTH: 505
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1101

Alignment Scores:
Pred. No.: 139 Length: 505
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-998-598-1101 (1-505)

QY 7 IleAspPheIleIlePhe 12
| | | | | | | | | | | | | | | |
Db 369 ATTGATTTTATTATTATTTT 386

RESULT 21
US-09-764-877-243
; Sequence 243, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 243
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (474)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (579)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-877-243

Alignment Scores:
Pred. No.: 159 Length: 579
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-764-877-243 (1-579)

QY 1 PheGlnAlaAsnCysGly 6
| | | | | | | | | | | | | | | |
Db 344 TTCCAAGCAAACTGCGGT 361

RESULT 22
US-09-864-761-8345
; Sequence 8345, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
```

```
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 8345
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC015971.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 8.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6
US-09-864-761-8345
```

```
Alignment Scores:
Pred. No.: 165 Length: 600
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0
```

US-09-854-133-587 (1-16) x US-09-864-761-8345 (1-600)

```
QY 6 GlyileaspPheIleIle 11
Db 489 GGCATTGACTTCATCATT 506
```

```
RESULT 23
US-09-974-300-6399/c
```

```
; Sequence 6399, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6399
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-6399
```

```
Alignment Scores:
Pred. No.: 191 Length: 696
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0
```

US-09-854-133-587 (1-16) x US-09-974-300-6399 (1-696)

```
QY 10 IleIlePheTrpIlePhe 15
|||||
```

```
Db 587 ATAATTTCTGGATTTT 570
```

## RESULT 24

```
US-09-867-550-1741
; Sequence 1741, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1741
; LENGTH: 989
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: Wherein n is one of a or t or c or g
US-09-867-550-1741
```

```
Alignment Scores:
Pred. No.: 269 Length: 989
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0
```

us-09-854-133-587.oligo.rnpb

Tue May 13 12:12:59 2003

```
US-09-854-133-587 (1-16) x US-09-867-550-1741 (1-989)
QY 11 IlePheTrpIlePheTrp 16
Db 557 ATATTTGGATCTTCTGG 574

RESULT 25
US-09-741-669-167/c
; Sequence 167, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; TITLE OF INVENTION: proliferation of E. coli
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 167
; LENGTH: 1188
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1188)
US-09-741-669-167

Alignment Scores:
Pred. No.: 321 Length: 1188
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservatave: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-741-669-167 (1-1188)
QY 6 GlyIleAspPheIleIle 11
Db 782 GGTATAGATTATTATAAT 765

RESULT 26
US-09-815-242-4288/c
; Sequence 4288, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8550
; LENGTH: 1263
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1263)
US-09-815-242-8550

Alignment Scores:
Pred. No.: 341 Length: 1263
Score: 6.00 Matches: 6

US-09-854-133-587 (1-16) x US-09-815-242-4288 (1-1260)
QY 7 IleAspPheIleIlePhe 12
Db 1191 ATCGATTTTATTATTTC 1174

RESULT 27
US-09-815-242-8550/c
; Sequence 8550, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8550
; LENGTH: 1263
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1263)
US-09-815-242-8550

Alignment Scores:
Pred. No.: 341 Length: 1263
Score: 6.00 Matches: 6
```

\_\_\_\_\_

Age Group	Percentage
18-24	10%
25-34	20%
35-44	30%
45-54	25%
55-64	15%
65-74	10%
75-84	5%
85+	5%



FILE REFERENCE: BGI-121CP2  
CURRENT APPLICATION NUMBER: US/09/746,660A  
CURRENT FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 09/606740  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 09/603124  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 60/141031  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: 60/142101  
PRIOR FILING DATE: 1999-07-02  
PRIOR APPLICATION NUMBER: 60/148613  
PRIOR FILING DATE: 1999-08-12  
PRIOR APPLICATION NUMBER: 60/187970  
PRIOR FILING DATE: 2000-03-09  
PRIOR APPLICATION NUMBER: DE 19931420.9  
PRIOR FILING DATE: 1999-07-08  
NUMBER OF SEQ ID NOS: 125  
SOFTWARE: PatentIn Vers. 2.0  
SEQ ID NO 11  
LENGTH: 1330  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (101)..(1330)  
OTHER INFORMATION: FRXA01009  
US-09-746-660A-11

Alignment Scores:  
Pred. No.: 359 Length: 1330  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-746-660A-11 (1-1330)

QY 8 AspPheillelePheTrp 13  
|||||  
Db 70 GACTTCATAATATTGG 87

RESULT 31  
US-10-142-231-43  
Sequence 43, Application US/10142231  
Publication No. US2003007796A1  
GENERAL INFORMATION:  
APPLICANT: Croteau, Rodney et al.  
TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES  
FILE REFERENCE: 62773  
CURRENT APPLICATION NUMBER: US/10/142,231  
CURRENT FILING DATE: 2002-05-08  
PRIOR APPLICATION NUMBER: 60/165,250  
PRIOR FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 43  
LENGTH: 1455  
TYPE: DNA  
ORGANISM: Taxus cuspidata  
US-10-142-231-43

Alignment Scores:  
Pred. No.: 391 Length: 1455  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-10-142-231-43 (1-1455)

QY 8 AspPheillelePheTrp 13  
|||||  
Db 602 GACTTCATCATCTTGG 619  
RESULT 32  
US-10-142-231-44  
Sequence 44, Application US/10142231  
Publication No. US2003007796A1  
GENERAL INFORMATION:  
APPLICANT: Croteau, Rodney et al.  
TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES  
FILE REFERENCE: 62773  
CURRENT APPLICATION NUMBER: US/10/142,231  
CURRENT FILING DATE: 2002-05-08  
PRIOR APPLICATION NUMBER: 60/165,250  
PRIOR FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 44  
LENGTH: 1455  
TYPE: DNA  
ORGANISM: Taxus cuspidata  
US-10-142-231-44

Alignment Scores:  
Pred. No.: 391 Length: 1455  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-10-142-231-44 (1-1455)

QY 8 AspPheillelePheTrp 13  
|||||  
Db 602 GACTTCATCATCTTGG 619

RESULT 33  
US-09-746-660A-9  
Sequence 9, Application US/09746660A  
Publication No. US20030049804A1  
GENERAL INFORMATION:  
APPLICANT: Pompejus, Markus  
APPLICANT: Kroger, Burkhard  
APPLICANT: Schroder, Hartwig  
APPLICANT: Zelder, Oskar  
APPLICANT: Haberhauer, Gregor  
APPLICANT: Kim, Jun-Won  
APPLICANT: Lee, Heung-Schick  
APPLICANT: Hwang, Byung-Joon  
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
METABOLIC PATHWAY PROTEINS  
FILE REFERENCE: BGI-121CP2  
CURRENT APPLICATION NUMBER: US/09/746,660A  
CURRENT FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 09/606740  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 09/603124  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 60/141031  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: 60/142101  
PRIOR FILING DATE: 1999-07-02  
PRIOR APPLICATION NUMBER: 60/148613  
PRIOR FILING DATE: 1999-08-12  
PRIOR APPLICATION NUMBER: 60/187970  
PRIOR FILING DATE: 2000-03-09  
PRIOR APPLICATION NUMBER: DE 19931420.9  
PRIOR FILING DATE: 1999-07-08  
NUMBER OF SEQ ID NOS: 125  
SOFTWARE: PatentIn Vers. 2.0  
SEQ ID NO 9



; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 3762  
; LENGTH: 2000  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-3762

Alignment Scores:  
Pred. No.: 534 Length: 2000  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-938-842A-3762 (1-2000)

QY 10 IleilePheTrpIlePhe 15  
|||||  
Db 1337 ATTATATTTGGATATT 1320

RESULT 38

US-09-938-842A-3983  
; Sequence 3983, Application US/09938842A  
; Patent No. US20020160378A1

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; FILE OF INVENTION: SAME, AND METHODS OF USE

; FILE REFERENCE: SCRIPI300-3

; CURRENT APPLICATION NUMBER: US/09/938,842A

; CURRENT FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227,866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264,647

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/300,111

; PRIOR FILING DATE: 2001-06-22

; NUMBER OF SEQ ID NOS: 5379

; SEQ ID NO 3983

; LENGTH: 2000

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-938-842A-3983

Alignment Scores:  
Pred. No.: 534 Length: 2000  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-938-842A-3983 (1-2000)

QY 9 PheIleilePheTrpIle 14  
|||||

Db 1609 TTTATATCTTTGGATT 1626

RESULT 39

US-09-938-842A-4984

; Sequence 4984, Application US/09938842A

; Patent No. US20020160378A1

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; FILE OF INVENTION: SAME, AND METHODS OF USE

; FILE REFERENCE: SCRIPI300-3

; CURRENT APPLICATION NUMBER: US/09/938,842A

; CURRENT FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227,866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264,647

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/300,111

; PRIOR FILING DATE: 2001-06-22

; NUMBER OF SEQ ID NOS: 5379

; SEQ ID NO 4984

; LENGTH: 2000

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-938-842A-4984

Alignment Scores:

Pred. No.: 534 Length: 2000  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-938-842A-4984 (1-2000)

QY 10 IleilePheTrpIlePhe 15  
|||||

Db 1546 ATTATATTTGGATATT 1563

RESULT 40

US-08-781-986A-360/c

; Sequence 360, Application US/08781986A

; Publication No. US20030054436A1

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 5255

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/781,986A

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Benson, Bob

; REGISTRATION NUMBER: 30,446

; REFERENCE/DOCKET NUMBER: PB248PP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 360:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2524 base pairs

```
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-360
Alignment Scores:
Pred. No.: 670 Length: 2524
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 7 Gaps: 0

US-09-854-133-587 (1-16) x US-08-781-986A-360 (1-2524)
QY 7 IleAspPheIleIlePhe 12
Db 1322 ATCGATTATTATTTTC 1305

RESULT 41
US-10-092-154-1254/c
; Sequence 1254, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1254
; LENGTH: 2692
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1254
Alignment Scores:
Pred. No.: 714 Length: 2692
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-10-092-154-1254 (1-2692)
QY 6 GlyIleAspPheIleIle 11
Db 2649 GGCATTGACTTCATTATA 2632

RESULT 42
US-09-764-847-1254/c
; Sequence 1254, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1254
; LENGTH: 2692
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1254
Alignment Scores:
Pred. No.: 714 Length: 2692
```

```
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-764-847-1254 (1-2692)
QY 6 GlyIleAspPheIleIle 11
Db 2649 GGCATTGACTTCATTATA 2632

RESULT 43
US-10-092-154-1255/c
; Sequence 1255, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1255
; LENGTH: 2694
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1255
Alignment Scores:
Pred. No.: 715 Length: 2694
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-10-092-154-1255 (1-2694)
QY 6 GlyIleAspPheIleIle 11
Db 2651 GGCATTGACTTCATTATA 2634

RESULT 44
US-09-764-847-1255/c
; Sequence 1255, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1255
; LENGTH: 2694
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1255
Alignment Scores:
Pred. No.: 715 Length: 2694
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-764-847-1255 (1-2694)
```



Tue May 13 12:12:59 2003

QY 6 GlyileAsppheille 11  
|||||  
Db 2651 GGCATTGACTTCATTATA 2634

```

RESULT 45
US-10-071-766-61
; Sequence 61, Application US/10071766
; Publication No. US20020192678A1
; GENERAL INFORMATION:
; APPLICANT: Huei-Mei Chen
; TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE
; FILE REFERENCE: PA-0043 US
; CURRENT APPLICATION NUMBER: US/10/071,766
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: PERL Program
; SEQ ID NO 61
; LENGTH: 2912
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020192678A1 037877.2
US-10-071-766-61

```

Alignment Scores:		
Pred. NO.:	771	2912
Score:	6.00	6
Percent Similarity:	100.00%	Length:
Best Local Similarity:	100.00%	Matches:
Query Match:	37.50%	Conservative:
DB:	9	Mismatches:
		Indels:
		Gaps:

US-09-854-133-587 (1-16) x US-10-071-766-61 (1-2912)

QY 10 ILEILEPHE<sup>TR</sup>PILEPHE 15  
DB 1275 ATTATATTTGGATCTTC 1292

Search completed: May 11, 2003, 18:46:09  
Job time : 33.9734 secs



Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 16:29:05 ; Search time 242.973 Seconds  
(without alignments)  
1066.486 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FQACGIDFIIFWIFW 16

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 32305347

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 60 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_174133\_718/app\_query.fasta\_1.462  
-DB-EST -QFMT=fastap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=60  
-DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=45 -MODE=LOCAL -OUTFMT=pt  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn\_1\_1\_2013\_erunat\_05052003\_174133\_718 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Result No.	Score	Match	Length	DB	ID	Description
C 1	8	50.0	876	12	BF693034	BF693034 602080030
C 2	7	43.8	240	9	AU073495	AU073495 AU073495
C 3	7	43.8	284	10	BB261269	BB261269 BB261269
C 4	7	43.8	336	12	BE954331	BE954331 UI-M-CEI-
C 5	7	43.8	438	12	BG446254	BG446254 GA_EB003
6	7	43.8	440	12	BG446243	BG446243 GA_EB003
7	7	43.8	450	17	AQ115066	AQ115066 CIT-HSP-2
C 8	7	43.8	452	9	AU270164	AU270164 AU270164
C 9	7	43.8	457	9	AU269237	AU269237 AU269237
10	7	43.8	460	12	BG543829	BG543829 E1546 Chi
C 11	7	43.8	465	17	AZ246141	AZ246141 RPCI-23-3
C 12	7	43.8	506	12	BF356741	BF356741 PM3-HT090
C 13	7	43.8	511	12	BG097509	BG097509 EST461944
C 14	7	43.8	529	13	BJ528250	BJ528250 BJ528250
C 15	7	43.8	572	12	BE971441	BE971441 601651432
C 16	7	43.8	598	12	BG571089	BG571089 602591696
C 17	7	43.8	612	17	BH539329	BH539329 BOGMZ37TR
C 18	7	43.8	617	17	AZ316177	AZ316177 1M0033A16
C 19	7	43.8	620	17	AZ937525	AZ937525 2M0195C17
C 20	7	43.8	642	14	BU008061	BU008061 QGH6F17.Y
C 21	7	43.8	675	14	BQ868209	BQ868209 QGD13F14.
C 22	7	43.8	710	9	AU136235	AU136235 AU136235
C 23	7	43.8	717	17	BH695614	BH695614 BOH2214TF
C 24	7	43.8	747	17	BH273281	BH273281 CH230-23H
C 25	7	43.8	753	17	BH422412	BH422412 BOHHD57TR
C 26	7	43.8	794	17	BH566426	BH566426 BOHSG71TR
C 27	7	43.8	799	17	AQ898772	AQ898772 HS_3130_A
C 28	7	43.8	864	17	BH152622	BH152622 ENTWG74TR
C 29	7	43.8	879	17	AZ692543	AZ692543 ENTW09TR
C 30	7	43.8	882	17	AZ692069	AZ692069 ENTME42TR
C 31	7	43.8	883	12	BF675705	BF675705 602083612
C 32	7	43.8	908	17	AZ545454	AZ545454 ENTGT31TR
C 33	7	43.8	908	17	BH130713	BH130713 ENTNX17TF
C 34	7	43.8	971	12	BF215221	BF215221 601846206
C 35	7	43.8	996	10	BE250032	BE250032 600943035
C 36	7	43.8	1077	13	BI260864	BI260864 602971086
C 37	6	37.5	91	14	RO1587	RO1587 ye75e11.r1
C 38	6	37.5	107	9	AL651945	AL651945 AL651945
C 39	6	37.5	135	9	AV060675	AV060675 AV060675
C 40	6	37.5	143	10	AW582814	AW582814 2fj3239c
C 41	6	37.5	144	12	BG189371	BG189371 RST8415.A
C 42	6	37.5	151	9	AA885939	AA885939 oj38a07.s
C 43	6	37.5	151	12	BE720567	BE720567 QV4-HT089
C 44	6	37.5	152	10	AV626699	AV626699 AV626699
C 45	6	37.5	152	10	AW794472	AW794472 RC6-UM001
C 46	6	37.5	152	10	BB073462	BB073462 BB073462
C 47	6	37.5	160	10	AW989089	AW989089 uf23a01.y
C 48	6	37.5	162	17	AZ357634	AZ357634 1M0099N15
C 49	6	37.5	167	10	BB422820	BB422820 BB422820
C 50	6	37.5	175	9	AA873237	AA873237 oh78e06.s
C 51	6	37.5	180	9	AV075220	AV075220 AV075220
C 52	6	37.5	181	9	AI152761	AI152761 ud94c08.r
C 53	6	37.5	183	10	BB575708	BB575708 BB575708
C 54	6	37.5	196	10	AV360305	AV360305 AV360305
C 55	6	37.5	199	9	AI156316	AI156316 ue09c09.r
C 56	6	37.5	199	9	AV087856	AV087856 AV087856
C 57	6	37.5	200	10	AW582852	AW582852 gh09c2b29
C 58	6	37.5	200	13	BI188910	BI188910 d4a11fs.r
C 59	6	37.5	200	17	AZ386368	AZ386368 1M0145K16
C 60	6	37.5	200	17	AZ716460	AZ716460 RPCI-24-1

ALIGNMENTS

## RESULT 1

BF693034/c  
 LOCUS 876 bp mRNA linear EST 22-DEC-2000  
 DEFINITION 602080030F1 NIH\_MGC\_81 Homo sapiens cDNA clone IMAGE:4244503 5', mRNA sequence.  
 ACCESSION BF693034  
 VERSION BF693034.1 GI:11978442  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 876)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: CLONETECH Laboratories, Inc.  
 CDNA Library Preparation: CLONETECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LLCMI056 row: 1 column: 08  
 High quality sequence stop: 595.  
 Location/Qualifiers

1. .876  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4244503"  
 /clone\_lib="NIH\_MGC\_81"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site\_1: SfiI (ggcgcctcgcc); Site\_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 355 a 164 c 175 g 182 t  
 ORIGIN

## FEATURES

source

Alignment Scores:

Pred. No.: 5.31 Length: 876

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 50.00% Indels: 0

DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BF693034 (1-876)

QY 8 AspPheIleIlePheTrpIlePhe 15

Db 572 GACTTCATTATATTTTGGATTTT 549

RESULT 2

AU073495/c

LOCUS 240 bp mRNA linear EST 24-JUN-1999

DEFINITION AU073495 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium

discoideum cDNA clone SSH359, mRNA sequence.

ACCESSION AU073495

VERSION AU073495.1 GI:5179916

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum

Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 284)

AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: [genome-res@gsc.riken.go.jp](mailto:genome-res@gsc.riken.go.jp)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 284)

AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: [genome-res@gsc.riken.go.jp](mailto:genome-res@gsc.riken.go.jp)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 284)

AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

## REFERENCE

1 (bases 1 to 240)

Drushihara, H.

Developmental cDNA in Dictyostelium discoideum (1999)

Unpublished (1999)

COMMENT Contact: Hideko Urushihara

Institute of Biological Sciences

University of Tsukuba

1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan

Tel: 81-298-53-4664

Fax: 81-298-53-6614

Email: [hideko@biol.tsukuba.ac.jp](mailto:hideko@biol.tsukuba.ac.jp)

PROJECT = Dictyostelium discoideum cDNA project in Japan.

Location/Qualifiers

1. .240

/organism="Dictyostelium discoideum"

/strain="AX4"

/db\_xref="taxon:44689"

/clone="SSH359"

/clone\_lib="Dictyostelium discoideum SS (H.Urushihara)"

/dev\_stage="slug"

BASE COUNT 151 a 16 c 14 g 59 t

ORIGIN

Alignment Scores:

Pred. No.: 28.1 Length: 240

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 43.75% Indels: 0

DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x AU073495 (1-240)

QY 9 PheIleIlePheTrpIlePhe 15

Db 50 TTTTATTATTTTCTGGATTTT 30

RESULT 3

BB261269/c

LOCUS 284 bp mRNA linear EST 07-JUL-2000

DEFINITION BB261269 RIKEN full-length enriched, 7 days neonate cerebellum Mus

musculus cDNA clone A730098B10 3' similar to AL049783 Novel human

gene mapping to chromosome 13, mRNA sequence.

ACCESSION BB261269

VERSION BB261269.1 GI:8957721

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 284)

AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: [genome-res@gsc.riken.go.jp](mailto:genome-res@gsc.riken.go.jp)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 284)

AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: [genome-res@gsc.riken.go.jp](mailto:genome-res@gsc.riken.go.jp)



AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mEST@mail.nih.gov  
Oligo-dt track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements The tissue for this library was contributed by Dr. Xin-Yuan Fu, Yale University School of Medicine  
Seq primer: M13 Forward  
POLYA=NO.

FEATURES  
source  
Location/Qualifiers  
1. .336  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-CE1-bak-h-05-0-UI"  
/clone\_lib="NIH\_BMAP\_Ret3\_N"  
/dev\_stage="6 weeks"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The NIH\_BMAP\_Ret3\_N library is a normalized library derived from mouse retina tissue. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu. The tissue for this library was contributed by Dr. Xin-Yuan Fu, Yale University School of Medicine  
TAG\_SEQ=None found"  
BASE COUNT 92 a 62 c 82 g 100 t  
ORIGIN

Alignment Scores:  
Pred. No.: 36.7 Length: 336  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 12 Gaps: 0  
US-09-854-133-587 (1-16) x BE954331 (1-336)  
QY 8 AspPheIleIlePheTrpIle 14  
|||||  
Db 49 GATTTCATCATATTCCTGGATA 29  
RESULT 5  
BG446254 438 bp mRNA linear EST 15-MAR-2001  
LOCUS GA\_Eb0033P23f Gossypium arboreum 7-10 dpa fiber library Gossypium  
DEFINITION arboreum cDNA clone GA\_Eb0033P23f, mRNA sequence.  
ACCESSION BG446254  
VERSION BG446254.1 GI:13355966  
KEYWORDS EST.  
SOURCE Gossypium arboreum.  
ORGANISM Gossypium arboreum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.  
REFERENCE 1 (bases 1 to 438)

URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki  
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh,M., Kitsuunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES  
source  
Location/Qualifiers  
1. .284  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="A730098B10"  
/clone\_lib="RIKEN full-length enriched, 7 days neonate cerebellum"  
/tissue\_type="cerebellum"  
/dev\_stage="7 days neonate"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATTCGAGTAAATTAATATCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."  
BASE COUNT 94 a 70 c 60 g 60 t  
ORIGIN

Alignment Scores:  
Pred. No.: 32.1 Length: 284  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 10 Gaps: 0  
US-09-854-133-587 (1-16) x BB261269 (1-284)  
QY 5 CysGlyIleAspPheIleIle 11  
|||||  
Db 188 TGTGGTATAGATTTCATCATC 168  
RESULT 4  
BE954331/c 336 bp mRNA linear EST 29-APR-2002  
LOCUS UI-M-CE1-bak-h-05-0-UI.s1 NIH\_BMAP\_Ret3\_N Mus musculus cDNA clone  
DEFINITION UI-M-CE1-bak-h-05-0-UI 3', mRNA sequence.  
ACCESSION BE954331  
VERSION BE954331.1 GI:10596750  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 336)

AUTHORS Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry  
D., Wood,T.C., Leslie,A. and Wilkins,T.A.  
TITLE An integrated analysis of the genetics, development, and evolution  
of the cotton fiber  
JOURNAL Unpublished (2000)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATACGACTCACTATAGGG  
High quality sequence start: 6  
High quality sequence stop: 438.

FEATURES source  
Location/Qualifiers  
1. .438  
/organism="Gossypium arboreum"  
/strain="AKA"  
/cultivar="8400"  
/db\_xref="taxon:29729"  
/clone="GA\_Eb0033P23f"  
/clone\_lib="Gossypium arboreum 7-10 dpa fiber library"  
/tissue\_type="Fibers isolated from bolls harvested 7-10  
dpa"  
/lab\_host="E. coli"  
/note="Vector: PBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"  
BASE COUNT 158 a 61 c 95 g 124 t  
ORIGIN

Alignment Scores:  
Pred. No.: 45.3 Length: 438  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 12 Gaps: 0  
US-09-854-133-587 (1-16) x BG446254 (1-438)  
QY 10 IlellePheTrpIlePheTrp 16  
|||||  
Db 308 ATTATATTTGGATATTTGG 328  
RESULT 6  
BG446243  
LOCUS  
DEFINITION GA\_Eb0033N21f Gossypium arboreum 7-10 dpa fiber library Gossypium  
arboreum cDNA clone GA\_Eb0033N21f, mRNA sequence.  
ACCESSION BG446243  
VERSION BG446243.1 GI:13355955  
KEYWORDS EST.  
SOURCE  
ORGANISM  
Gossypium arboreum.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.  
1 (bases 1 to 440)  
Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry  
D., Wood,T.C., Leslie,A. and Wilkins,T.A.  
TITLE An integrated analysis of the genetics, development, and evolution  
of the cotton fiber  
JOURNAL Unpublished (2000)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATACGACTCACTATAGGG  
High quality sequence start: 12  
High quality sequence stop: 419.

FEATURES source  
Location/Qualifiers  
1. .440  
/organism="Gossypium arboreum"  
/strain="AKA"  
/cultivar="8400"  
/db\_xref="taxon:29729"  
/clone="GA\_Eb0033N21f"  
/clone\_lib="Gossypium arboreum 7-10 dpa fiber library"  
/tissue\_type="Fibers isolated from bolls harvested 7-10  
dpa"  
/lab\_host="E. coli"  
/note="Vector: PBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"  
BASE COUNT 144 a 76 c 96 g 124 t  
ORIGIN

Alignment Scores:  
Pred. No.: 45.5 Length: 440  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 12 Gaps: 0  
US-09-854-133-587 (1-16) x BG446243 (1-440)  
QY 10 IlellePheTrpIlePheTrp 16  
|||||  
Db 310 ATTATATTTGGATATTTGG 330  
RESULT 7  
AQ115066  
LOCUS  
DEFINITION CIT-HSP-2374E12.TF CIT-HSP Homo sapiens genomic clone 2374E12, DNA  
sequence.  
ACCESSION AQ115066  
VERSION AQ115066.1 GI:3491187  
KEYWORDS GSS.  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 450)  
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,  
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and  
Venter,J.C.  
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready  
Map Building  
JOURNAL Unpublished (1998)  
COMMENT Other\_GSSs: CIT-HSP-2374E12.TR  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadamsetigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13-21  
Class: BAC ends.

FEATURES source  
Location/Qualifiers  
1. .450  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2374E12"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2:  
HindIII"  
BASE COUNT 120 a 62 c 71 g 196 t  
ORIGIN

SOURCE Dictyostelium discoideum.  
 ORGANISM Dictyostelium discoideum  
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
 REFERENCE 1 (bases 1 to 457)  
 AUTHORS Urushihara,H., Morio,T., Saito,T., Koriki,E., Ochiai,H., Maeda,M., Takeuchi,I., Kohara,Y. and Tanaka,Y.  
 TITLE Population analysis of cDNAs from unicellular and multicellular stages of Dictyostelium discoideum  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Hideko Urushihara  
 Institute of Biological Sciences  
 University of Tsukuba  
 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan  
 Tel: 81-298-53-4664  
 Fax: 81-298-53-6614  
 Email: hideko@biol.tsukuba.ac.jp.

FEATURES  
 Location/Qualifiers  
 1..457  
 /organism="Dictyostelium discoideum"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /clone="VSI791"  
 /clone\_lib="VS"  
 /sex="mat A"  
 /dev\_stage="vegetative"  
 187 a 69 c 71 g 129 t 1 others  
 BASE COUNT  
 ORIGIN

Alignment Scores:  
 Pred. No.: 46.9 Length: 457  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x AU269237 (1-457)  
 QY 9 PheillePheTrpIlePhe 15  
 |||||  
 Db 309 TTTATAATTTTGGATCTTT 289

RESULT 10  
 BG543829 460 bp mRNA linear EST 01-MAY-2002  
 LOCUS E1546 Chinese cabbage etiolated seedling library Brassica rapa  
 DEFINITION subsp. pekinensis cDNA clone E1546, mRNA sequence.  
 ACCESSION BG543829  
 VERSION BG543829.1 GI:20374809  
 KEYWORDS EST.  
 SOURCE Brassica rapa subsp. pekinensis.  
 ORGANISM Brassica rapa subsp. pekinensis  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
 REFERENCE 1 (bases 1 to 460)  
 AUTHORS Ryu,S.H., Kang,J.S., Kang,C.-h., Kim,C.Y., Choi,Y.J., Lee,S.-H., Bahk,J.D., Lee,S.Y., Cho,M.J. and Lim,C.O.  
 TITLE Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Lim, C.O.  
 Plant Molecular Biology & Biotechnology Research Centre  
 Gyeongsang National University  
 #900 Gazwa-dong, Jinju 660-701, Korea  
 Tel: 82 55 751 6255  
 Fax: 82 55 759 9363  
 Email: colim@ngae.gsnu.ac.kr  
 Seq primer: T7  
 Location/Qualifiers  
 1..460  
 /organism="Brassica rapa subsp. pekinensis"  
 /cultivar="Jangwon"  
 /db\_xref="taxon:51351"

Alignment Scores:  
 Pred. No.: 46.3 Length: 450  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x AQ115066 (1-450)  
 QY 9 PheillePheTrpIlePhe 15  
 |||||  
 Db 424 TTTATCATATTTGGATATTT 444

RESULT 8  
 AU270164/c  
 LOCUS AU270164 VS Dictyostelium discoideum cDNA clone VSI588 3', mRNA  
 DEFINITION sequence.  
 ACCESSION AU270164  
 VERSION AU270164.1 GI:20528962  
 KEYWORDS EST.  
 SOURCE Dictyostelium discoideum.  
 ORGANISM Dictyostelium discoideum  
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
 REFERENCE 1 (bases 1 to 452)  
 AUTHORS Urushihara,H., Morio,T., Saito,T., Koriki,E., Ochiai,H., Maeda,M., Takeuchi,I., Kohara,Y. and Tanaka,Y.  
 TITLE Population analysis of cDNAs from unicellular and multicellular stages of Dictyostelium discoideum  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Hideko Urushihara  
 Institute of Biological Sciences  
 University of Tsukuba  
 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan  
 Tel: 81-298-53-4664  
 Fax: 81-298-53-6614  
 Email: hideko@biol.tsukuba.ac.jp.

FEATURES  
 Location/Qualifiers  
 1..452  
 /organism="Dictyostelium discoideum"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /clone="VSI588"  
 /clone\_lib="VS"  
 /sex="mat A"  
 /dev\_stage="vegetative"  
 184 a 60 c 78 g 130 t  
 BASE COUNT  
 ORIGIN

Alignment Scores:  
 Pred. No.: 46.5 Length: 452  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x AU270164 (1-452)  
 QY 9 PheillePheTrpIlePhe 15  
 |||||  
 Db 164 TTTATAATTTTGGATCTTT 144

RESULT 9  
 AU269237/c  
 LOCUS AU269237 VS Dictyostelium discoideum cDNA clone VSI791 3', mRNA  
 DEFINITION sequence.  
 ACCESSION AU269237  
 VERSION AU269237.1 GI:20528035  
 KEYWORDS EST.

```
/clone="E1546"
/clone_lib="Chinese cabbage etiolated seedling library"
/tissue_type="Etiolated seedling"
/lab_host="XL-1 Blue"
/note="Vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"

BASE COUNT      137 a      93 c      87 g      143 t
ORIGIN

Alignment Scores:
Pred. No.:      47.2      Length:      460
Score:          7.00      Matches:      7
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      43.75%      Indels:      0
DB:              12      Gaps:      0

US-09-854-133-587 (1-16) x BG543829 (1-460)

QY      9 PheillelePheTrpIlePhe 15
Db      127 TTTATAATCTTTGGATT 147

RESULT 11
AZ246141/c
LOCUS
DEFINITION
  RPCI-23-38C13.TJB RPCI-23 Mus musculus genomic clone RPCI-23-38C13,
  DNA sequence.
ACCESSION
  AZ246141
VERSION
  AZ246141.1 GI:8559432
KEYWORDS
  GSS.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
REFERENCE
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 465)
AUTHORS
  Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
  ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
  and Fraser,C.M.
  Mouse BAC End Sequences from Library RPCI-23
  Unpublished (1999)
  Other_GSSs: RPCI-23-38C13.TV
  Contact: Shaying Zhao
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: szhao@tigr.org
  Clones are derived from the mouse BAC library RPCI-23. For BAC
  library availability, please contact Pieter de Jong
  (pieter@dejong.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
  or from Resea ch Genetics (info@resgen.com). BAC end page:
  http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
  Plate: 38 row: C column: 13
  Seq primer: SP6
  Class: BAC ends.
FEATURES
  Location/Qualifiers
    1..465
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="RPCI-23-38C13"
      /clone_lib="RPCI-23"
      /sex="Female"
      /lab_host="DH10B"
      /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
  EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
  brain genomic DNA was isolated and partially digested
  with a combination of EcoRI and EcoRI Methylase. Size
  selected DNA was cloned into the pBACe3.6 vector at the
  EcoRI sites. The ligation products were transformed into
  DH10B electrocompetent cells (BRL Life Technologies)."
```

```
BASE COUNT      236 a      60 c      65 g      104 t
ORIGIN

Alignment Scores:
Pred. No.:      47.6      Length:      465
Score:          7.00      Matches:      7
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      43.75%      Indels:      0
DB:              17      Gaps:      0

US-09-854-133-587 (1-16) x AZ246141 (1-465)

QY      9 PheillelePheTrpIlePhe 15
Db      294 TTTATAATCTTTGGATT 274

RESULT 12
BF356741/c
LOCUS
DEFINITION
  PM3-HT0909-130900-002-h06 HT0909 Homo sapiens cDNA, mRNA sequence.
ACCESSION
  BF356741
VERSION
  BF356741.1 GI:11315815
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
REFERENCE
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  1 (bases 1 to 506)
AUTHORS
  Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
  Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
  Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
  Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
  ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
  Simpson,A.J.
  Shotgun sequencing of the human transcriptome with ORF expressed
  sequence tags
  Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  20202663
  Contact: Simpson A.J.G.
  Laboratory of Cancer Genetics
  Ludwig Institute for Cancer Research
  Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
  Brazil
  Tel: +55-11-2704922
  Fax: +55-11-2707001
  Email: asimpson@ludwig.org.br
  This sequence was derived from the FAPESP/LICR Human Cancer Genome
  Project. This entry can be seen in the following URL
  (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&t2=PM3-HT0909-
  130900-002-h06&t3=2000-09-13&t4=1)
  Seq primer: puc 18 forward
  High quality sequence start: 12
  High quality sequence stop: 505.
FEATURES
  Location/Qualifiers
    1..506
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="HT0909"
      /dev_stage="Adult"
      /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
  Site_2: SmaI; A mini-library was made by cloning products
  derived from ORESTES PCR (U.S. Letters Patent application
  No. 196,716 - Ludwig Institute for Cancer Research)
  profiles into the pUC 18 vector. Reverse transcription of
  tissue mRNA and cDNA amplification were performed under
  low stringency conditions."
```

```
BASE COUNT      165 a      98 c      106 g      137 t
ORIGIN

Alignment Scores:
Pred. No.:      50.9      Length:      506
Score:          7.00      Matches:      7
```



Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BF356741 (1-506)

QY 6 GlyIleAspPheIleIlePhe 12  
|||||  
Db 300 GGTATTGATTTTATCATTTTC 280

RESULT 13

BG097509

LOCUS BG097509 511 bp mRNA linear EST 29-JAN-2001  
DEFINITION EST461944 potato leaves and petioles Solanum tuberosum cDNA clone  
CSTB47G13 5' sequence, mRNA sequence.

ACCESSION BG097509

VERSION BG097509.1 GI:12587460

KEYWORDS EST.

SOURCE potato.

ORGANISM Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE 1 (bases 1 to 511)

AUTHORS van der Hoeven,R.S., Bezzerides,J., Holt,I.E., Liang,F., Cho,J.,  
Utterback,T., Hansen,C.L., Doan,B., Bougri,O., Buell,C.R., Ronning  
,C.M., Fry,W.E., Tanksley,S.D. and Baker,B.

TITLE Generation of ESTs from potato leaves and petioles

JOURNAL Unpublished (2000)

COMMENT

Contact: Cathy Ronning

The Institute for Genomic Research

For clone info: please contact Research Genetics, Libraries

Division tel 1-800-711-6195, email cdna@resgen.com.

FEATURES

source

1. .511  
/organism="Solanum tuberosum"  
/cultivar="Kennebec"  
/db\_xref="taxon:4113"  
/clone="CSTB47G13"  
/clone\_lib="potato leaves and petioles"  
/tissue\_type="leaflets and petioles"  
/dev\_stage="8 weeks old plants"  
/lab\_host="SOLR"

/note="Vector: pBlueScript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; Tissue was supplied by Dr. Fry (Cornell University).  
Leaflets and petioles were isolated from 8 week old  
greenhouse grown plants. The plants were watered and  
fertilized freely. The tissue was immediately frozen in  
liquid nitrogen."

BASE COUNT 153 a 64 c 112 g 182 t  
ORIGIN

Alignment Scores:

Pred. No.: 51.3 Length: 511  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BG097509 (1-511)

QY 3 AlaAsnCysGlyIleAspPhe 9

|||||

Db 420 GCAATGTGGTATTGATTTT 440

RESULT 14

BJ528250/c

LOCUS BJ528250 529 bp mRNA linear EST 09-AUG-2002

DEFINITION BJ528250 MF01SSB cDNA Oryzias latipes cDNA clone MF01SSB032A10 5',

mRNA sequence.

ACCESSION BJ528250

VERSION BJ528250.1

GI:22187062

KEYWORDS EST.

SOURCE Japanese medaka.

ORGANISM

Oryzias latipes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

REFERENCE 1 (bases 1 to 529)

AUTHORS Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.

TITLE Medaka EST Project in Takeda's lab

JOURNAL Unpublished (2001)

COMMENT

Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

FEATURES

source

1. .529  
/organism="Oryzias latipes"  
/strain="Hd-rR"  
/db\_xref="taxon:8090"  
/clone="MF01SSB032A10"  
/clone\_lib="MF01SSB cDNA"  
/sex="mixture of female and male"  
/tissue\_type="whole embryo"  
/dev\_stage="segmentation stage 20 - 25"

BASE COUNT 147 a 83 c 83 g 214 t 2 others  
ORIGIN

Alignment Scores:

Pred. No.: 52.7 Length: 529  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 13 Gaps: 0

US-09-854-133-587 (1-16) x BJ528250 (1-529)

QY 9 PheIleIlePheTrpIlePhe 15

|||||

Db 488 TTTATCATATTTTGGATTTT 468

RESULT 15

BE971441/c

LOCUS BE971441

DEFINITION BE971441 NIH\_MGC\_81 Homo sapiens cDNA clone IMAGE:3935070 5',

mRNA sequence.

ACCESSION BE971441

VERSION BE971441.1 GI:10584777

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 572)

AUTHORS NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: CLONETECH Laboratories, Inc.

cDNA Library Preparation: CLONETECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LICM776 row: 0 column: 07

High quality sequence stop: 571.

## FEATURES

## source

Location/Qualifiers  
1. .572  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3935070"  
/clone\_lib="NIH\_MGC\_81"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site\_1: SfiI (ggccgctcgcc); Site\_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:  
5'-ATTCTAGAGCGGCGGCACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 235 a 109 c 104 g 124 t  
ORIGIN

## Alignment Scores:

Pred. No.: 56.1 Length: 572  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BE971441 (1-572)

QY 9 Pheillelephetrpillephe 15

Db 513 TTCATTATATTTGGATTTT 493

## RESULT 16

BG571089/c

## LOCUS

DEFINITION BG571089 598 bp mRNA linear EST 10-APR-2001  
602591696F1 NIH\_MGC\_79 Homo sapiens cDNA clone IMAGE:4714247 5',  
mRNA sequence.

ACCESSION BG571089

VERSION BG571089.1 GI:13578742

## KEYWORDS

SOURCE EST.

ORGANISM human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 598)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)

Tissue Procurement: CLONTECH Laboratories, Inc.

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCMI558 row: p column: 24

High quality sequence stop: 598.

## FEATURES

## source

Location/Qualifiers  
1. .598  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4714247"  
/clone\_lib="NIH\_MGC\_79"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: placenta; Vector: pDNR-LIB (Clontech);  
Site\_1: SfiI (ggccgctcgcc); Site\_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCGGCGGCACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

BASE COUNT 188 a 107 c 144 g 159 t  
ORIGIN

## Alignment Scores:

Pred. No.: 58.1 Length: 598  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BG571089 (1-598)

QY 6 GlylleAsppheillelephe 12

Db 28 GGGATTGACTTCATCATTTT 8

## RESULT 17

BH539329/c

## LOCUS

DEFINITION BH539329 612 bp DNA linear GSS 14-DEC-2001  
BOGMZ37TR BOGM Brassica oleracea genomic clone BOGMZ37, DNA

sequence.

ACCESSION BH539329

VERSION BH539329.1 GI:17782077

## KEYWORDS

SOURCE GSS.

ORGANISM Brassica oleracea.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 612)

Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other\_GSSs: BOGMZ37TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: [cdtown@tigr.org](mailto:cdtown@tigr.org)

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

## FEATURES

## source

Location/Qualifiers  
1. .612  
/organism="Brassica oleracea"  
/strain="T01000DH3"  
/db\_xref="taxon:3712"  
/clone="BOGMZ37"  
/clone\_lib="BOGM"  
/note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 195 a 116 c 150 g 151 t  
ORIGIN

## Alignment Scores:

Pred. No.: 59.2 Length: 612  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x BH539329 (1-612)



## Alignment Scores:

Pred. No.: 59.8 Length: 620  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x AZ937525 (1-620)

QY 2 GlnAlaAsnCysGlyIleAsp 8  
|||||  
Db 395 CAAGCAATGTGGCATTGAT 375

RESULT 20  
BU008061

LOCUS

DEFINITION QGH6F17.yg.ab1 QG\_EFGHJ lettuce serriola Lactuca sativa cDNA clone  
QGH6F17, mRNA sequence. EST 22-AUG-2002

ACCESSION BU008061

VERSION BU008061.1 GI:22442456

KEYWORDS

SOURCE

ORGANISM

Lactuca sativa.  
Lactuca sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridaceae; euasterids II; Asterales; Asteraceae; Lactuceae;  
Lactuca.

REFERENCE

AUTHORS

1 (bases 1 to 642)

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,  
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison  
, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,  
Church, S., Jackson, L. and Bradford, K.  
Lectuue and Sunflower ESTs from the Compositae Genome Project  
<http://compgenomics.ucdavis.edu/>  
Unpublished (2002)

TITLE

JOURNAL

COMMENT

Contact: Alexander Kozik [R.W.Michelmore]  
Department of Vegetable Crops, R.W.Michelmore Lab  
University of California at Davis (UCD)  
Asmundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-742-1742  
Fax: 1-(530)-752-9659

Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]  
belongs to contig QG\_CA\_Contig3097, see <http://cgpdb.ucdavis.edu/>  
for details.

Plate: QGH6 row: F column: 17.

FEATURES

source

Location/Qualifiers  
1..642  
/organism="Lactuca sativa"  
/cultivar="L.serriola"  
/db\_xref="taxon:4236"  
/clone="QGH6F17"  
/clone\_lib="QG\_EFGHJ lettuce serriola"  
/lab\_host="E.coli"  
/note="Vector: pBRCNASfiAB; The library was constructed  
from 10 different sources of RNA from a single genotype.  
Separate cDNAs were generated using primers that  
incorporated unique 5' and 3' tags to distinguish each  
source of RNA. cDNAs were then pooled, size-fractionated,  
directionally cloned into a custom medium-copy vector and  
transformations made with four size classes to minimize  
size bias. Details of each source of RNA and library  
construction can be obtained at <http://cgpdb.ucdavis.edu/>  
TAG\_LIB=QG\_EFGHJ lettuce serriola  
TAG\_TISSUE=germinating seeds  
TAG\_SEQ=TCTGTGCGGG"

BASE COUNT 201 a 113 c 162 g 165 t 1 others  
ORIGIN

## Alignment Scores:

Pred. No.: 61.5 Length: 642  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 14 Gaps: 0

US-09-854-133-587 (1-16) x BU008061 (1-642)

QY 6 GlyIleAspPheIleIlePhe 12  
|||||  
Db 444 GGGATTGATTATCATCTTC 464

RESULT 21

BQ868209

LOCUS

DEFINITION

BQ868209 QGD13F14.yg.ab1 QG\_ABCDI lettuce salinas Lactuca sativa cDNA clone  
QGD13F14, mRNA sequence. EST 14-AUG-2002

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Lactuca sativa.

Lactuca sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridaceae; euasterids II; Asterales; Asteraceae; Lactuceae;  
Lactuca.

REFERENCE

AUTHORS

1 (bases 1 to 675)

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,  
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison  
, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,  
Church, S., Jackson, L. and Bradford, K.  
Lectuue and Sunflower ESTs from the Compositae Genome Project  
<http://compgenomics.ucdavis.edu/>  
Unpublished (2002)

TITLE

JOURNAL

COMMENT

Contact: Alexander Kozik [R.W.Michelmore]  
Department of Vegetable Crops, R.W.Michelmore Lab  
University of California at Davis (UCD)  
Asmundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-742-1742  
Fax: 1-(530)-752-9659

Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]  
belongs to contig QG\_CA\_Contig3097, see <http://cgpdb.ucdavis.edu/>  
for details.

Plate: QGD13 row: F column: 14.

FEATURES

source

Location/Qualifiers  
1..675  
/organism="Lactuca sativa"  
/cultivar="Salinas"  
/db\_xref="taxon:4236"  
/clone="QGD13F14"  
/clone\_lib="QG\_ABCDI lettuce salinas"  
/lab\_host="E.coli"  
/note="Vector: pBRCNASfiAB; The library was constructed  
from 10 different sources of RNA from a single genotype.  
Separate cDNAs were generated using primers that  
incorporated unique 5' and 3' tags to distinguish each  
source of RNA. cDNAs were then pooled, size-fractionated,  
directionally cloned into a custom medium-copy vector and  
transformations made with four size classes to minimize  
size bias. Details of each source of RNA and library  
construction can be obtained at <http://cgpdb.ucdavis.edu/>  
TAG\_LIB=QG\_ABCDI lettuce salinas  
TAG\_TISSUE=germinating seeds  
TAG\_SEQ=TCTGTGCGGG"

BASE COUNT 215 a 116 c 172 g 172 t  
ORIGIN

## Alignment Scores:

Pred. No.: 64 Length: 675  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 14 Gaps: 0



US-09-854-133-587 (1-16) x BQ868209 (1-675)

QY 6 GlyIleAspPheIleIlePhe 12  
|||||

Db 438 GGGATTGATTTTATCACTTC 458

## RESULT 22

AU136235

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS

Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,  
Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and  
Isogai, T.

HRI human cDNA project

Unpublished (2000)

Contact: Takao Isogai

Genomics Laboratory

Helix Research Institute

1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

Tel: 81-438-52-3975

Fax: 81-438-52-3986

Email: genomics@hri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix  
Research Institute; cDNA library construction; Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

## FEATURES

source

1. 710

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="PLACE1003825"

/clone\_lib="PLACE1"

/tissue\_type="placenta"

/note="Vector: pME18SFL3"

BASE COUNT 206 a 150 c 104 g 244 t 6 others

ORIGIN

## Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-854-133-587 (1-16) x AU136235 (1-710)

QY 1 PheGlnAlaAsnCysGlyIle 7

|||||

Db 506 TTTCAGCCCAACTGTGGGATA 526

## RESULT 23

BH695614

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Brassica oleracea

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other\_GSSs: BOHZZ14TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seg primer: TF

Class: sheared ends.

## FEATURES

source

1. 717

/organism="Brassica oleracea"

/strain="TOL000DH3"

/db\_xref="taxon:3712"

/clone="BOHZZ14"

/clone\_lib="BO\_2\_3\_KB"

/note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared  
genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-854-133-587 (1-16) x BH695614 (1-717)

QY 6 GlyIleAspPheIleIlePhe 12

|||||

Db 472 GGTATAGATTATATATATTT 492

## RESULT 24

BH273281

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 747)

Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn

A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de

Jong, P. and Fraser, C.M.

Rat BAC End Sequences from Library CHORI-230 EcoRI segment

Unpublished (1999)

Other\_GSSs: CH230-23H8.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(http://www.chori.org/bacpac/rat230.htm). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources

(http://www.chori.org/bacpac/or

ering\_information.htm). BAC end

page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html

Plate: 23 row: H column: 8

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1. .747

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db\_xref="taxon:10116"

/clone="CH230-23H8"

/clone\_lib="CHORI-230 Segment 1"

/sex="Female"

/cell\_type="Brain"

/note="Vector: pTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI;

CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by

Pieter de Jong"

192 a 178 c 156 g 221 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 69.4 Length: 747

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 43.75% Indels: 0

DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x BH273281 (1-747)

QY 9 PheillePheTrpPhe 15

|||||

Db 390 TTTATCATTTCTCTGGATCTTT 410

RESULT 25

BH422412

LOCUS

DEFINITION BOHHD57R BOHH Brassica oleracea genomic clone BOHHD57, DNA

753 bp DNA linear GSS 12-DEC-2001

sequence.

ACCESSION BH422412

VERSION BH422412.1 GI:17608140

KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 753)

Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other\_GSSs: BOHHD57TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. .753

/organism="Brassica oleracea"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

/clone="BOHHD57"

/clone\_lib="BOHH"

/note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared

genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 69.9 Length: 753

Score:

Percent Similarity: 7.00

Best Local Similarity: 100.00%

Query Match: 43.75%

DB: 17

Matches: 7

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-09-854-133-587 (1-16) x BH422412 (1-753)

QY 9 PheillePheTrpPhe 15

|||||

Db 354 TTTATAATCTCTCTGGATTTT 374

RESULT 26

BH566426/c

LOCUS

DEFINITION

BH566426

BOHSG71TR BOHS Brassica oleracea genomic clone BOHSG71, DNA

sequence.

ACCESSION BH566426

VERSION BH566426.1 GI:17818266

KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 794)

Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other\_GSSs: BOHSG71TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. .794

/organism="Brassica oleracea"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

/clone="BOHSG71"

/clone\_lib="BOHS"

/note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared

genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 297 a 102 c 109 g 286 t

ORIGIN

Alignment Scores:

Pred. No.: 72.9 Length: 794

Score: 7.00 Matches: 7

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 43.75%

DB: 17

US-09-854-133-587 (1-16) x BH566426 (1-794)

QY 10 IleillePheTrpPheTrp 16

|||||

Db 586 ATAATTTCTCGATTTTGG 566

RESULT 27

AQ898772

LOCUS

DEFINITION

AQ898772

HS\_3130\_A2\_C05\_T7C CIT Approved Human Genomic Sperm Library D Homo

sapiens genomic clone Plate-3130 Col-10 Row-E, DNA sequence.

ACCESSION AQ898772

VERSION AQ898772.1 GI:6354962

KEYWORDS GSS.

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 799)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 3130 row: E column: 10  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 799.

FEATURES  
source

1. 799  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=3130 Col=10 Row=E"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in  
E-Coli DH10B"  
BASE COUNT 233 a 171 c 141 g 253 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 73.2 Length: 799  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x AQ898772 (1-799)

QY 10 IleIlePheTrpIlePheTrp 16  
|||||  
Db 498 ATTATCTTCTGGATATTTGG 518

RESULT 28  
BH152622  
LOCUS BH152622 864 bp DNA linear GSS 27-AUG-2001  
DEFINITION ENTPG74TR Entamoeba histolytica Sheared DNA Entamoeba histolytica  
genomic, DNA sequence.  
ACCESSION BH152622  
VERSION BH152622.1 GI:15316561  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica  
Eukaryota; Entamoebidae; Entamoeba.  
REFERENCE 1 (bases 1 to 864)  
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
HM1:IMSS sheared DNA library (2001)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543

Email: bjloftus@tigr.org  
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared  
DNA library  
Seq primer: M13-Reverse  
Class: shotgun  
High quality sequence start: 20  
High quality sequence stop: 829.

FEATURES  
source

1. 864  
/organism="Entamoeba histolytica"  
/strain="HM1:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pHOS1; Site\_1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + 1 method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barell, Oxford University Press, 1999)."  
BASE COUNT 328 a 102 c 115 g 319 t  
ORIGIN

Alignment Scores:  
Pred. No.: 78 Length: 864  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x BH152622 (1-864)

QY 1 PheGlnAlaAsnCysGlyIle 7  
|||||  
Db 84 TTTCAGGCTAATTGCGGTATT 104

RESULT 29  
AZ692543

LOCUS AZ692543 879 bp DNA linear GSS 14-DEC-2000  
DEFINITION ENTIW09TR Entamoeba histolytica Sheared DNA Entamoeba histolytica  
genomic, DNA sequence.  
ACCESSION AZ692543  
VERSION AZ692543.1 GI:11829689  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica  
Eukaryota; Entamoebidae; Entamoeba.  
REFERENCE 1 (bases 1 to 879)  
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
HM1:IMSS sheared DNA library  
JOURNAL Unpublished (2000)  
COMMENT Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@tigr.org  
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared  
DNA library  
Seq primer: M13-Reverse  
Class: shotgun  
High quality sequence start: 15  
High quality sequence stop: 754.

FEATURES  
Location/Qualifiers

## source

1. .879  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pHOS1; Site\_1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barell, Oxford University Press, 1999)."

BASE COUNT 314 a 110 c 130 g 325 t  
ORIGIN

## Alignment Scores:

Pred. No.: 79 Length: 879  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x AZ692543 (1-879)

QY 1 PheGlnAlaAsnCysGlyIle 7  
|||||  
Db 509 TTTCAGGCTAATTGCGGTATT 529

## RESULT 30

AZ692069

LOCUS

DEFINITION ENTME42TR Entamoeba histolytica Sheared DNA linear GSS 14-DEC-2000  
genomic, DNA sequence.

ACCESSION AZ692069

VERSION AZ692069.1 GI:11829110

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica

REFERENCE 1 (bases 1 to 882)

AUTHORS Loftus,B., Van Aken,S. and Fraser,C.

TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMSS sheared DNA library

COMMENT Unpublished (2000)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjloftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Seq primer: M13-Reverse

Class: shotgun

High quality sequence start: 19

High quality sequence stop: 824.

Location/Qualifiers

## FEATURES

source

1. .882  
/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db\_xref="taxon:5759"

/clone\_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHOS1; Site\_1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barell, Oxford University Press, 1999)."

BASE COUNT 293 a 134 c 124 g 331 t  
ORIGIN

## Alignment Scores:

Pred. No.: 79.2 Length: 882  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x AZ692069 (1-882)

QY 9 PheillellePheTrpIlePhe 15  
|||||

Db 804 TTTCATTATATTGGATTTT 824

## RESULT 31

BF675705/c

LOCUS

DEFINITION 602083612F1 NIH\_MGC\_83 Homo sapiens cDNA clone linear EST 21-DEC-2000  
mRNA sequence.

ACCESSION BF675705

VERSION BF675705.1 GI:11949600

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 883)

TITLE NIH-MGC http://mgc.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: CLONETECH Laboratories, Inc.

CDNA Library Preparation: CLONETECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM1065 row: h column: 04

High quality sequence stop: 608.

Location/Qualifiers

## FEATURES

source

1. .883  
/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4247859"

/clone\_lib="NIH\_MGC\_83"

/lab\_host="DH10B (Tl phage-resistant)"

/note="Organ: prostate; Vector: pDNR-LIB (Clontech);

Site\_1: SfiI (ggccgctcgcc); Site\_2: SfiI (ggccattatggcc

); 5' and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGGCCATTATGGC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCCGAGCGCCGACATG-dT(30)BN-3'

(where B = A, C, or G and N = A, C, G, or T). Average

insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA)."

BASE COUNT 336 a 205 c 164 g 178 t

ORIGIN



Alignment Scores:		
Pred. No.:	79.3	883
Score:	7.00	7
Percent Similarity:	100.00%	0
Best Local Similarity:	100.00%	0
Query Match:	43.75%	0
DB:	12	0
		Length:
		Matches:
		Conservative:
		Mismatches:
		Indels:
		Gaps:

US-09-854-133-587 (1-16) x BF675705 (1-883)

Qy 9 pheillelphetrpilephe 15  
|||  
Db 573 TTCATTATATTTGGATTTTT 553

RESULT	32
AZ545454	
LOCUS	AZ545454 linear GSS 14-NOV-2000
DEFINITION	ENTGT31TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION	AZ545454
VERSION	AZ545454.1 GI:11166245
KEYWORDS	GSS.
SOURCE	Entamoeba histolytica.
ORGANISM	Entamoeba histolytica Eukaryota; Entamoebidae; Entamoeba. 1 (bases 1 to 908)
REFERENCE	Lofthus,B., Van Aken,S. and Fraser,C. Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library
JOURNAL	Unpublished (2000)
COMMENT	Contact: Brendan J Loftus

BASE COUNT	370 a	97 c	149 g	292 t
ORIGIN				
Alignment Scores:				
Pred. No.:	81.1		Length:	908
Score:	7.00		Matches:	7
Percent Similarity:	100.00%		Conservative:	0
Best Local Similarity:	100.00%		Mismatches:	0
Query Match:	43.75%		Indels:	0
DB:	17		Gaps:	0

US-09-854-133-587 (1-16) x A2545454 (1-908)

Qy 9 pheillelePhetrpilephe 15  
 |||||  
 pb 658 TTCATAATATTGGATCTTT 678

RESULT 33  
BH130713/c

BH130713	908 bp	DNA	linear	GSS 07-AUG-2001
ENTNX17TF	Entamoeba histolytica	Sheared DNA	Entamoeba histolytica	
genomic, DNA sequence.				
BH130713				
BH130713.1	GI:15089062			
GSS.				
Entamoeba histolytica.				
Entamoeba histolytica				
Entamoebidae; Entamoeba.				
Eukaryota; Entamoebidae; Entamoeba.				
1 (bases 1 to 908)				
Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.				
Determination of clone end sequences from Entamoeba histolytica				
HM1:IMSS sheared DNA library (2001)				
Unpublished (2001)				
Contact: Brendan J Loftus				

BASE COUNT	352 a	116 c	152 g	288 t
ORIGIN				
Alignment Scores:				
Pred. No.:	81.1		Length:	908
Score:	7.00		Matches:	7
Percent Similarity:	100.00%		Conservative:	0
Best Local Similarity:	100.00%		Mismatches:	0
Query Match:	43.75%		Indels:	0
DB:	17		Gaps:	0

US-09-854-133-587 (1-16) X BH130713 (1-908)

QY 9 pheIleIlepheTrpIlephe 15  
|||||  
772 TTtATtATtTTCTGCATATTC 752

RESULT 34  
BF215221/C

LOCUS BF215221 971 bp mRNA linear EST 06-NOV-2000  
DEFINITION 601846206F1 NIH\_MGC\_55 Homo sapiens cDNA clone IMAGE:4077432 5',  
mRNA sequence.  
ACCESSION BF215221  
VERSION BF215221.1 GI:11108807  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 971)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONETECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM932 row: k column: 01  
High quality sequence start: 19  
High quality sequence stop: 62.  
Location/Qualifiers  
1. 971  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4077432"  
/clone\_lib="NIH\_MGC\_55"  
/tissue\_type="from acute myelogenous leukemia"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);  
Site\_1: SfiI (ggcgctcgcc); Site\_2: SfiI (ggccattatggcc  
); Double-stranded cDNA was prepared from cell line RNA.  
5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGGCCATTATGCC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCGCGGCGGCATG-dt(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.65 kb (range 0.9-4.0 kb). 14/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."  
BASE COUNT 329 a 201 c 207 g 234 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 85.6 Length: 971  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 12 Gaps: 0  
US-09-854-133-587 (1-16) x BF215221 (1-971)  
QY 9 PheillePheTrpIlePhe 15  
|||||  
Db 657 TTCATTATATTGGATTTT 637  
RESULT 35  
BE250032  
LOCUS BE250032 996 bp mRNA linear EST 13-JUL-2000  
DEFINITION 600943035F1 NIH\_MGC\_15 Homo sapiens cDNA clone IMAGE:2959497 5',  
mRNA sequence.  
ACCESSION BE250032  
VERSION BE250032.1 GI:9120137  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 996)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLCM51 row: f column: 10  
High quality sequence stop: 669.  
Location/Qualifiers  
1. 996  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2959497"  
/clone\_lib="NIH\_MGC\_15"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)"  
BASE COUNT 196 a 319 c 284 g 197 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 87.3 Length: 996  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 10 Gaps: 0  
US-09-854-133-587 (1-16) x BE250032 (1-996)  
QY 1 PheGlnAlaAsnCysGlyIle 7  
|||||  
Db 665 TTCCAGCGGAATTGTGGATC 685  
RESULT 36  
BI260864  
LOCUS BI260864 1077 bp mRNA linear EST 17-JUL-2001  
DEFINITION 602971086F1 NIH\_MGC\_12 Homo sapiens cDNA clone IMAGE:5110436 5',  
mRNA sequence.  
ACCESSION BI260864  
VERSION BI260864.1 GI:14819571  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1077)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Incyte Genomics, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: L1AM11267 row: 1 column: 21  
High quality sequence stop: 700.

FEATURES  
source

Location/Qualifiers  
1. .1077  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5110436"  
/clone\_lib="NIH\_MGC\_12"  
/tissue\_type="cervical carcinoma cell line"  
/lab\_host="DH10B"  
/note="Organ: cervix; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: Sali; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.4 kb. Library prepared by Life  
Technologies."

BASE COUNT 335 a 258 c 210 g 274 t  
ORIGIN

Alignment Scores: 92.9 Length: 1077  
Pred. No.: 7.00 Matches: 7  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 43.75% Indels: 0  
Query Match: 13 Gaps: 0  
DB:

US-09-854-133-587 (1-16) x BI260864 (1-1077)

QY 9 PheillelePheTrpIlePhe 15  
|||||  
Db 138 TTCATAATCTTTGGATTTC 158

RESULT 37

R01587 91 bp mRNA linear EST 31-MAR-1995  
LOCUS ye75ell.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone  
DEFINITION IMAGE:123596 5' similar to gb:J03779 NEPRILYSIN (HUMAN);, mRNA  
sequence.

ACCESSION R01587  
VERSION R01587.1 GI:751323  
KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 91)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman  
, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
Rifkin, L., Rohlif, T., Soares, M., Tan, F., Trevaskis, E., Waterston  
, R., Williamson, A., Wohldmann, P. and Wilson, R.

TITLE

JOURNAL The WashU-Merck EST Project  
COMMENT Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1458

High quality sequence stops: 70 Source: IMAGE Consortium, LLNL This  
clone is available royalty-free through LLNL; contact the IMAGE  
Consortium (info@image.llnl.gov) for further information.

Insert Length: 1458 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 70.

FEATURES

Location/Qualifiers  
1. .91  
/organism="Homo sapiens"  
/db\_xref="GDB:476141"  
/db\_xref="taxon:9606"  
/clone="IMAGE:123596"  
/clone\_lib="Soares fetal liver spleen 1NFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"

/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' AACTGGAAGAAATTAATAAGATCTTTTCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 29 a 16 c 11 g 33 t  
ORIGIN

Alignment Scores: 192 Length: 91  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 37.50% Indels: 0  
Query Match: 14 Gaps: 0  
DB:

US-09-854-133-587 (1-16) x R01587 (1-91)

QY 1 PheGlnAlaAsnCysGly 6  
|||||  
Db 67 TTTCAAGCCAACTGTGGA 84

RESULT 38

AL651945 107 bp mRNA linear EST 13-DEC-2001  
LOCUS AL651945 XGC-gastrula Silurana tropicalis cDNA clone TGas046111 5',  
DEFINITION mRNA sequence.

ACCESSION AL651945  
VERSION AL651945.1 GI:17662266  
KEYWORDS EST.  
SOURCE western clawed frog.  
ORGANISM Silurana tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
Xenopodinae; Silurana.  
1 (bases 1 to 107)

REFERENCE

AUTHORS Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.  
TITLE Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Huckle E

Sanger Centre  
Hinxtun, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: TGas046111.sp6  
Sequencing primer: SP6

This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Aaron M. Zorn.

FEATURES

Location/Qualifiers  
1. .107  
/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone="TGas046111"  
/clone\_lib="XGC-gastrula"  
/dev\_stage="gastrula (stages 10.5-13 mixed)"  
/lab\_host="Escherichia coli XL1-blue"  
/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA  
was oligo dT primed from 5ug of poly A+ RNA from stages  
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated  
into pCS107 with EcoRI at the 5' end and NotI at the 3'  
end."

BASE COUNT 42 a 18 c 13 g 34 t  
ORIGIN

Alignment Scores:

Pred. No.: 219 Length: 107  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x AL651945 (1-107)

QY 9 PheilleillePheTrpIle 14  
|||||  
Db 85 TTTATTATTTCTGGATC 102

RESULT 39  
AV060675

LOCUS

DEFINITION AV060675 Mus musculus pancreas C57BL/6J adult Mus musculus CDNA  
clone 1810064A19, mRNA sequence.

ACCESSION AV060675

VERSION AV060675.1 GI:5160422

KEYWORDS

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 135)

REFERENCE

AUTHORS Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,  
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara,  
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,  
Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H., Owa, C.,  
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara,  
Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y., Tominaga, N.,  
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,  
Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

TITLE

JOURNAL

COMMENT

Unpublished (1999)

Contact: Chie Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp

Thermotabilization and thermoactivation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length CDNA  
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

FEATURES

source

Location/Qualifiers

1..135

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="1810064A19"

/clone\_lib="Mus musculus pancreas C57BL/6J adult"

/sex="male"

/tissue\_type="pancreas"

/dev\_stage="adult"

/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia  
) with a modified polylinker; Site\_1: Not I; Site\_2: Eco  
RI; 1st strand CDNA was primed with a Not I - oligo(dT)  
primer [5,  
TGTTACCAATCTGAAGTGGGAGCGCGCGAATGTTTGTGTGTGTGTGTGTGTGTGT  
T 3']; double-stranded CDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."

BASE COUNT

ORIGIN

37 a 40 c 26 g 32 t

Alignment Scores:

Pred. No.: 263

Length:

135

Score:

Percent Similarity: 6.00

Best Local Similarity: 100.00%

Query Match: 100.00%

DB: 37.50%

Indels: 0

Gaps: 0

Mismatches: 6

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-09-854-133-587 (1-16) x AV060675 (1-135)

QY 8 AspPheilleillePheTrp 13

|||||

Db 61 GACTTCATCATTTTGG 78

RESULT 40

AW582814

LOCUS

DEFINITION

AW582814 2fj323gc Neuronal Differentiation of the NT2/D1 cell line. Homo

sapiens cDNA 3' similar to EST, mRNA sequence.

ACCESSION AW582814

VERSION AW582814.1 GI:7382060

KEYWORDS

SOURCE human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 143)

Bevort, M.

Analysis of gene expression during neuronal differentiation of

NT2/D1 cells

Unpublished (2000)

Contact: Bevort M

Department of Growth and Reproduction GR-5064

Copenhagen University Hospital

Blegdamsvej 9, 2100 Copenhagen, Denmark

Tel: +45 35455081

Fax: +45 35456054

Email: maja@biobase.dk

The EST is up regulated, during neuronal differentiation of the

NT2/D1 cell line (replated fully differentiated neurones not

analysed).

PCR Primers

FORWARD: GAGCAGGAATTGC

BACKWARD: AAGCTTTTCTTTTGTG

Seq primer: T7, CY5-TAATACGACTCACTATAGGGCC

High quality sequence stop: 143.

Location/Qualifiers

1..143

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Neuronal Differentiation of the NT2/D1 cell

line."

/cell\_line="NT2/D1"

/note="The EST is derived from direct sequencing of a

Differential Display fragment. Laboratory manuals are

available from <http://www.biobase.dk/-ddbase>"

BASE COUNT 42 a 30 c 25 g 43 t 3 others

ORIGIN

Alignment Scores:

Pred. No.: 276

Score: 6.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 37.50%

DB: 10

Length: 143

Mismatches: 6

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-09-854-133-587 (1-16) x AW582814 (1-143)

QY 6 GlyIleAspPheIleIle 11

|||||

Db 24 GGCATTGACTTCATTATA 41

RESULT 41

BG189371/c



LOCUS BG189371 144 bp mRNA linear EST 21-APR-2001  
DEFINITION RST8415 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG189371  
VERSION BG189371.1 GI:13711058  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 144)  
AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,  
Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J.,  
Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith  
, E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher  
, J., Danzig, J. and Ducar, M.  
TITLE Creation of genome-wide protein expression libraries using random  
activation of gene expression  
JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)  
MEDLINE 21227151  
COMMENT Contact: Scott J. Cain  
Athersys, Inc.  
3201 Carnegie Ave, Cleveland, OH 44115, USA  
Tel: 216 431 9900  
Fax: 216 361 9596  
Email: scain@atersys.com  
High quality sequence stop: 144.  
FEATURES  
source  
1..144  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Athersys RAGE Library"  
/cell\_line="HT1080"  
/note="See 'Creation of Genome-wide Protein Expression  
Libraries using Random Activation of Gene Expression',  
Nature Biotechnology, in press. Note that even though the  
cell type indicated is HT1080, since a random activation  
method was used, these sequence tags are not necessarily  
expressed in HT1080 under normal circumstances."  
BASE COUNT 54 a 21 c 19 g 50 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 277 Length: 144  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 12 Gaps: 0  
US-09-854-133-587 (1-16) x BG189371 (1-144)  
QY 4 AsnCysGlyIleAspPhe 9  
|||||  
Db 129 AATTGCGGAATAGATTTT 112  
RESULT 42  
AA885939  
LOCUS AA885939 151 bp mRNA linear EST 09-JUN-1998  
DEFINITION oj38a07.s1 NCI\_CGAP\_Kid3 Homo sapiens cDNA clone IMAGE:1500564 3',  
mRNA sequence.  
ACCESSION AA885939  
VERSION AA885939.1 GI:3001047  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 151)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert length: 1241 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 59.  
FEATURES  
source  
1..151  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:1500564"  
/clone\_lib="NCI\_CGAP\_Kid3"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer,  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified pT7T3 vector. mRNA  
source: 2 pooled kidneys. Library went through one round  
of normalization. Library constructed by Bento Soares and  
M. Fatima Bonaldo."  
BASE COUNT 48 a 30 c 28 g 45 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 288 Length: 151  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 9 Gaps: 0  
US-09-854-133-587 (1-16) x AA885939 (1-151)  
QY 6 GlyIleAspPheIleIle 11  
|||||  
Db 37 GGCATTGACTTCATTATA 54  
RESULT 43  
BE720567/c  
LOCUS BE720567 151 bp mRNA linear EST 12-SEP-2000  
DEFINITION QV4-HT0894-270700-318-e03 HT0894 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE720567  
VERSION BE720567.1 GI:10108832  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 151)  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL  
MEDLINE  
COMMENT

sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=QV4-HT0894-270  
700-318-e03&t3=2000-07-27&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 6  
High quality sequence stop: 151.  
Location/Qualifiers  
1. .151  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0894"  
/dev\_stage="Adult"  
/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
47 a 31 c 32 g 41 t

BASE COUNT  
ORIGIN

Alignment Scores:  
Pred. No.: 288 Length: 151  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BE720567 (1-151)

QY 6 GlyIleAspPheIleIle 11  
|||||

Db 143 GGCATTGACTTCATCATT 126

RESULT 44  
AV626699/c

LOCUS  
DEFINITION  
AV626699 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas  
reinhardtii cDNA clone LCL013h08\_r 5', mRNA sequence.

ACCESSION  
VERSION  
AV626699.1 GI:10788979

KEYWORDS  
SOURCE  
EST.  
ORGANISM  
Chlamydomonas reinhardtii.  
Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadaceae; Chlamydomonas.

REFERENCE  
1 (bases 1 to 152)

AUTHORS  
Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,  
Nakamura, Y. and Tabata, S.  
Generation of expressed sequence tags from low-CO2 and high-CO2  
adapted cells of Chlamydomonas reinhardtii  
DNA Res. 7 (5), 305-307 (2000)

TITLE  
Chlamydomonas reinhardtii.

JOURNAL  
MEDLINE  
COMMENT  
Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
source

1. .152  
/organism="Chlamydomonas reinhardtii"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone\_lib="LCL013h08\_r"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI; The cDNA library was constructed from cells cultured  
in a carbon stress acclimatized condition in which carbon  
dioxide concentration in the bubbling gas was changed from  
5% to 0.04%"

BASE COUNT  
ORIGIN

Alignment Scores:  
Pred. No.: 290 Length: 152  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x AV626699 (1-152)

QY 3 AlaAsnCysGlyIleAsp 8  
|||||

Db 39 GCGAACTGTGGTATCGAC 22

RESULT 45  
AW794472

LOCUS  
DEFINITION  
RC6-UM0014-210200-011-F04 UM0014 Homo sapiens cDNA, mRNA sequence.

ACCESSION  
VERSION  
AW794472.1 GI:7846342

KEYWORDS  
SOURCE  
EST.  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 152)

AUTHORS  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL  
MEDLINE  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=RC6-UM0014-210  
200-011-F04&t3=2000-02-21&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 152.  
Location/Qualifiers  
1. .152  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="UM0014"  
/dev\_stage="Adult"  
/note="Organ: uterus; Vector: puc18; Site\_1: SmaI; Site\_2:  
SmaI; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196

,716 - Ludwig Institute for Cancer Research) profiles  
into the pUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions." 19 t

BASE COUNT 51 a 33 c 49 g 19 t  
ORIGIN

Alignment Scores: 290 Length: 152  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 37.50% Indels: 0  
Query Match: 10 Gaps: 0  
DB:

US-09-854-133-587 (1-16) x AW794472 (1-152)

QY 3 AlaAsnCysGlyIleAsp 8

Db 118 GCAAACTGTGGCATAGAC 135

Search completed: May 11, 2003, 17:46:22  
Job time : 249.973 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 19:32:42 ; Search time 9.34513 Seconds  
(without alignments)  
228.141 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : A_Geneseq_101002.*			
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*		
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	100.0	16	AAE13851	Human T cell epitope
2	6	37.5	399	AAW05523	HCMV Towne strain
3	5	31.2	44	ABB38939	Peptide #6445 enco
4	5	31.2	44	ABB23925	Protein #5924 enco
5	5	31.2	44	AAW59589	Human brain expres
6	5	31.2	44	AAW72162	Human bone marrow
7	5	31.2	44	AAW19471	Peptide #5905 enco
8	5	31.2	44	AAW32422	Peptide #6459 enco
9	5	31.2	44	ABG41976	Human peptide enco
10	5	31.2	62	AAW30672	Amino acid sequenc

11	5	31.2	62	21	AAW28390	Human CD28 transme
12	5	31.2	62	22	AAW4314	Amino acid sequenc
13	5	31.2	62	22	AAW98753	Human CD28 extrace
14	5	31.2	62	22	AAW98800	Human CD28 extrace
15	5	31.2	62	22	AAW28678	Protein encoded by
16	5	31.2	62	23	AAU98026	Human IgG1 hinge/C
17	5	31.2	67	22	AAW92938	Human digestive sy
18	5	31.2	67	22	AAU20034	Human liver associ
19	5	31.2	67	23	ABP40895	Human liver antigen
20	5	31.2	68	21	AAW58985	Breast and ovarian
21	5	31.2	69	17	AAW94651	Mouse Fas-associat
22	5	31.2	72	21	AAW00588	Human secreted pro
23	5	31.2	74	23	ABP40821	Staphylococcus epi
24	5	31.2	76	22	ABG27612	Novel human diagno
25	5	31.2	81	18	AAW28161	Amino acid sequenc
26	5	31.2	82	22	AAW94589	Human reproductive
27	5	31.2	87	20	AAW60427	Human normal bladd
28	5	31.2	92	22	ABG27579	Novel human diagno
29	5	31.2	92	22	AAW90743	Human immune/haema
30	5	31.2	103	20	AAW24470	Human CD28 gene pr
31	5	31.2	104	22	AAW94419	Human reproductive
32	5	31.2	111	22	ABG30130	Novel human diagno
33	5	31.2	112	21	AAW40079	Arabidopsis thalia
34	5	31.2	118	23	ABB09650	Amino acid sequenc
35	5	31.2	120	23	ABB48192	Listeria monocytog
36	5	31.2	134	23	ABP40387	Staphylococcus epi
37	5	31.2	146	7	AAW60068	Sequence of HTLV-I
38	5	31.2	148	22	AAW92462	Human protein sequ
39	5	31.2	149	19	AAW85902	S. pneumoniae deri
40	5	31.2	150	18	AAW20434	H. pylori cytoplas
41	5	31.2	151	22	ABG30352	Novel human diagno
42	5	31.2	152	22	ABG22986	Novel human diagno
43	5	31.2	157	23	ABP29925	Streptococcus poly
44	5	31.2	157	23	ABP30763	Streptococcus poly
45	5	31.2	160	22	ABG12807	Novel human diagno
46	5	31.2	166	22	AAU71909	C. glutamicum meta
47	5	31.2	166	22	AAU71910	C. glutamicum meta
48	5	31.2	166	22	AAW90777	C glutamicum prote
49	5	31.2	166	22	AAW79741	Corynebacterium gl
50	5	31.2	166	22	AAW79742	Corynebacterium gl
51	5	31.2	166	22	AAW79886	Corynebacterium gl
52	5	31.2	166	22	AAW80068	Corynebacterium gl
53	5	31.2	166	22	AAW80069	Corynebacterium gl
54	5	31.2	170	8	AAW70262	Beta-glucuronidase
55	5	31.2	170	9	AAW80057	Beta-glucuronidase
56	5	31.2	195	15	AAW56557	Protein product of
57	5	31.2	195	23	ABW54469	Lactococcus lactis
58	5	31.2	197	21	AAW75378	Neisseria meningit
59	5	31.2	197	21	AAW75376	Neisseria gonorrhoe
60	5	31.2	198	21	AAW75377	Neisseria meningit
61	5	31.2	198	22	AAU25706	G protein-coupled
62	5	31.2	198	23	AAE21334	Mouse MrgC9 (mas-r
63	5	31.2	202	23	ABW89451	Human polypeptide
64	5	31.2	205	22	AAW78783	Human protein SEQ
65	5	31.2	206	23	AAE21332	Mouse MrgC7 (mas-r

ALIGNMENTS

RESULT 1	
AAE13851	
ID	AAE13851 standard; peptide; 16 AA.
XX	
AC	AAE13851;
XX	
DT	26-FEB-2002 (first entry)
XX	
DE	Human T cell epitope related to lung tumour-specific protein.
XX	
KW	Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW	antisense-therapy; vaccine; immune response; lung cancer;
KW	T cell epitope.

XX Homo sapiens.  
OS WO200172295-A2.  
XX  
PN  
XX  
XX  
PD 04-OCT-2001.  
XX  
XX  
PF 28-MAR-2001; 2001WO-US09991.  
XX  
XX  
PR 29-MAR-2000; 2000US-0538037.  
PR 05-JUN-2000; 2000US-0588937.  
PR 18-AUG-2000; 2000US-0640878.  
PR 22-SEP-2000; 2000US-234517P.  
PR 01-NOV-2000; 2000US-0704512.  
PR 14-DEC-2000; 2000US-0738973.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;  
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
XX  
XX WPI; 2001-639201/73.  
XX  
XX  
PT New human lung-specific polynucleotides and polypeptides for the  
XX diagnosis and treatment of disease e.g. lung cancer -  
XX  
XX Claim 2; Page 378; 378pp; English.  
XX  
XX The invention relates to isolated lung tumour-specific proteins and  
XX their corresponding cDNA molecules. Lung tumour-specific proteins and  
XX their antigen-presenting cells are useful for stimulating and/or  
XX expanding T cells specific for a tumour protein, and for inhibiting  
XX the development of cancer. The invention also relates to a composition  
XX useful for stimulating an immune response, and for treating cancer. The  
XX lung tumour specific oligonucleotide is useful in gene therapy and for  
XX diagnosis, detection and treatment of lung cancer. The present sequence  
XX is human T cell epitope related to lung tumour-specific protein.  
XX  
SQ Sequence 16 AA;  
  
Query Match 100.0%; Score 16; DB 22; Length 16;  
Best Local Similarity 100.0%; Pred. No. 9.8e-12;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 FQANCGIDFIIFWIFW 16  
Db 1 FQANCGIDFIIFWIFW 16  
  
RESULT 2  
AAW05523  
ID AAW05523 standard; Protein; 399 AA.  
XX  
XX AAW05523;  
XX  
XX 15-JAN-1997 (first entry)  
XX  
XX HCMV Towne strain UL154 protein.  
XX  
XX CMV; HCMV; vaccine; diagnosis; UL154.  
XX  
XX Human cytomegalovirus strain Towne (ATCC VR 977).  
XX  
XX WO9630387-A1.  
XX  
XX 03-OCT-1996.  
XX  
XX 26-MAR-1996; 96WO-US04100.  
XX  
XX 31-MAR-1995; 95US-0414926.  
XX  
XX (AVIR-) AVIRON.

PI Cha T, Spaete R;  
XX  
DR WPI; 1996-455265/45.  
DR N-PSDB; AAT41419.  
XX  
PT New isolated human cytomegalovirus nucleic acid - from Towne and  
PT Toledo strains, used to develop prods. for the diagnosis, prevention  
PT and treatment of human CMV infections  
XX  
PS Claim 5; Page 47-48; 150pp; English.  
XX  
CC Novel protein UL154 (AAW05523) is the product of an open reading  
CC frame found in a novel nucleic acid (AAT41419) isolated from the  
CC Towne strain of human cytomegalovirus (HCMV). UL154 and other  
CC novel proteins (see also AAW05521-22 and AAW05524) of the Towne strain,  
CC as well as novel and known proteins (see also AAW05500-20) from HCMV  
CC Toledo, can be produced in transformed host cells and used in the  
CC prodn. of subunit vaccines against HCMV. They may be surface  
CC glycoproteins that are immunogenic or responsible for tissue  
CC tropism, or may influence the immune response of an infected  
CC individual.  
XX  
SQ Sequence 399 AA;  
  
Query Match 37.5%; Score 6; DB 17; Length 399;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 9 FIIFWI 14  
Db 20 FIIFWI 25  
  
RESULT 3  
ABB38939  
ID ABB38939 standard; Peptide; 44 AA.  
XX  
XX ABB38939;  
XX  
XX 04-FEB-2002 (first entry)  
XX  
XX Peptide #6445 encoded by human foetal liver single exon probe.  
XX  
XX Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX  
XX Homo sapiens.  
XX  
XX WO200157277-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00669.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human fetal liver -  
XX  
XX Claim 27; SEQ ID NO 31574; 639pp + sequence listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 44 AA;  
  
Query Match 31.2%; Score 5; DB 22; Length 44;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 6 GIDFI 10  
| | | | |  
Db 9 GIDFI 13  
  
RESULT 4  
ABB23925  
ID ABB23925 standard; Protein; 44 AA.  
XX  
AC ABB23925;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Protein #5924 encoded by probe for measuring heart cell gene expression.  
XX  
KW Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00666.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
PS Claim 15; SEQ ID NO 25695; 530pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 44 AA;  
  
Query Match 31.2%; Score 5; DB 22; Length 44;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 6 GIDFI 10  
| | | | |  
Db 9 GIDFI 13  
  
RESULT 5  
AAM59589  
ID AAM59589 standard; Protein; 44 AA.  
XX  
AC AAM59589;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 31694.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00667.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483446/52.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
XX  
PS Example 4; SEQ ID NO: 31694; 650pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention.  
XX  
SQ Sequence 44 AA;  
  
Query Match 31.2%; Score 5; DB 22; Length 44;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 6 GIDFI 10  
| | | | |  
Db 9 GIDFI 13  
  
RESULT 6

```
AAM72162
ID AAM72162 standard; Protein; 44 AA.
XX AC AAM72162;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 32468.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX DR Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX PS Example 4; SEQ ID NO: 32468; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX SQ Sequence 44 AA;
XX Query Match 31.2%; Score 5; DB 22; Length 44;
XX Best Local Similarity 100.0%; Pred. No. 72;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 GIDFI 10
Db 9 GIDFI 13
XX RESULT 7
XX AAM19471
XX ID AAM19471 standard; Protein; 44 AA.
XX AC AAM19471;
XX DT 12-OCT-2001 (first entry)
XX DE Peptide #5905 encoded by probe for measuring cervical gene expression.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer.
XX OS Homo sapiens.
XX PN WO200157278-A2.
```

```
XX 09-AUG-2001.
PD 30-JAN-2001; 2001WO-US00670.
XX 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX DR Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX PS Claim 27; SEQ ID No 24297; 487pp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX (SENPs: see AAI10068-AAI28459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer.
XX CC Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 44 AA;
XX Query Match 31.2%; Score 5; DB 22; Length 44;
XX Best Local Similarity 100.0%; Pred. No. 72;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 GIDFI 10
Db 9 GIDFI 13
XX RESULT 8
XX AAM32422
XX ID AAM32422 standard; Protein; 44 AA.
XX AC AAM32422;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #6459 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
```



PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488897/53.  
DR  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human placenta -  
XX  
PS Claim 27; SEQ ID No 32691; 654pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes (SENP;  
CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders.  
XX  
SQ Sequence 44 AA;  
Query Match 31.2%; Score 5; DB 22; Length 44;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 GIDFI 10  
Db |||||  
9 GIDFI 13  
RESULT 9  
ABG41976  
ID ABG41976 standard; Peptide; 44 AA.  
XX  
AC ABG41976;  
XX  
DT 19-AUG-2002 (first entry)  
XX  
DE Human peptide encoded by genome-derived single exon probe SEQ ID 31641.  
XX  
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200186003-A2.  
XX  
PD 15-NOV-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00665.  
XX  
PR 04-FEB-2000; 2000US-180312P.  
PR 26-MAY-2000; 2000US-207456P.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-234687P.  
PR 27-SEP-2000; 2000US-236359P.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
DR

XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples -  
XX  
PS Claim 27; SEQ ID No 31641; 634pp; English.  
XX  
CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of  
CC probes; the novel set of probes which hybridise at high stringency to a  
CC nucleic acid expressed in the human lung; measuring gene expression in a  
CC sample derived from human lung, comprising (a) contacting the array with  
CC a collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of  
CC the array; identifying exons in a eukaryotic genome, comprising  
CC (a) algorithmically predicting at least one exon from genomic sequences  
CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene  
CC expression analysis, and for identifying exons in a gene, particularly  
CC using human lung derived mRNA and for the study of lung diseases  
CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
CC and hyaline membrane disease. The present sequence is a peptide/protein  
CC encoded by a single exon probe of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence . 44 AA;  
Query Match 31.2%; Score 5; DB 23; Length 44;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 GIDFI 10  
Db |||||  
9 GIDFI 13  
RESULT 10  
AAB30672  
ID AAB30672 standard; Protein; 62 AA.  
XX  
AC AAB30672;  
XX  
DT 19-MAR-2001 (first entry)  
XX  
DE Amino acid sequence of a signalling component cassette.  
XX  
KW Membrane-associated protein; transmembrane region; infectious disease;  
KW HIV infection; inflammatory disease; autoimmunity; rheumatoid arthritis;  
KW osteoarthritis; inflammatory bowel disease; cancer; allergic disease;  
KW atopic disease; asthma; eczema; congenital disease; cystic fibrosis;  
KW sickle cell anaemia; dermatologic disease; psoriasis; neurologic disease;

KW multiple sclerosis; organ transplant rejection; diabetes; CD28;  
KW graft-versus-host disease.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
PN WO200063374-A1.  
XX  
PD 26-OCT-2000.  
XX  
PF 17-APR-2000; 2000WO-GB01476.  
XX  
PR 16-APR-1999; 99GB-0008816.  
PR 16-APR-1999; 99GB-0008818.  
XX  
PA (CLLT ) CELLTech THERAPEUTICS LTD.  
PI Lawson ADG, Finney HM;  
XX  
DR WPI; 2000-687177/67.  
DR N-PSDB; AAC62383.  
XX  
PT Novel nucleic acid useful for treatment or prevention of disease such  
PT as cancer, infections, asthma, neurological disorder in humans, encodes  
PT a membrane-associated protein comprising synthetic transmembrane region  
PT  
XX  
PS Example 1; Fig 2; 42pp; English.  
XX  
CC The specification describes nucleic acids encoding a synthetic  
CC transmembrane region and membrane-associated proteins comprising a  
CC synthetic transmembrane region. Membrane-associated polypeptides and  
CC polynucleotides are useful in therapy and for preparing a medicament  
CC for the treatment or prevention of disease in humans. Diseases or  
CC disorders which may be treated include infectious diseases e.g. human  
CC immunodeficiency virus (HIV) infection, inflammatory disease/autoimmunity  
CC e.g. rheumatoid arthritis, osteoarthritis, inflammatory bowel disease,  
CC cancer, allergic/atopic disease e.g. asthma, eczema, congenital  
CC e.g. cystic fibrosis, sickle cell anaemia, dermatologic e.g. psoriasis,  
CC neurologic e.g. multiple sclerosis, transplants e.g. organ transplant  
CC rejection, graft-versus-host disease, metabolic/idiopathic disease  
CC e.g. diabetes. The present sequence represents a signalling component  
CC cassette, and comprises a human CD28 transmembrane region. The cassette  
CC is used to facilitate construction of chimeric receptors with modified  
CC or synthetic transmembrane regions.  
XX  
SQ Sequence 62 AA;  
  
Query Match 31.2%; Score 5; DB 21; Length 62;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 9 FIIFW 13  
Db 53 FIIFW 57  
  
RESULT 11  
AAB28390  
ID AAB28390 standard; Protein; 62 AA.  
XX  
AC AAB28390;  
XX  
DT 19-FEB-2001 (first entry)  
XX  
DE Human CD28 transmembrane region fragment.  
XX  
KW Oligonucleotide library; sequence block; human;  
KW CD28 transmembrane region.  
XX  
OS Homo sapiens.  
XX  
PN WO200063360-A1.

XX 26-OCT-2000.  
XX  
PF 17-APR-2000; 2000WO-GB01498.  
XX  
PR 16-APR-1999; 99GB-0008814.  
XX  
PA (CLLT ) CELLTech THERAPEUTICS LTD.  
XX  
PI Finney HM, Lawson ADG;  
XX  
DR WPI; 2000-679595/66.  
DR N-PSDB; AAC67772.  
XX  
PT Generating a library of DNA molecules of varying length and sequence in  
PT a desired orientation comprises ligating a mixture of double-stranded  
PT DNA molecules, and cutting ligated molecules in undesired orientations  
PT  
XX  
PS Example 1; Fig 3; 50pp; English.  
XX  
CC The present invention relates to a method for generating a library of  
CC oligonucleotides of varying length and sequence in a desired orientation.  
CC The method comprises ligating a double-stranded oligonucleotide mixture  
CC having 5' and 3' ends compatible to each other and corresponding to  
CC cleavage products of different but compatible restriction enzymes, and  
CC cutting ligated oligonucleotide, such that only molecules that are  
CC ligated in undesired orientations are cut, to produce a sequence block.  
CC The present sequence is a fragment of human CD28 transmembrane region.  
CC The coding sequence for the present protein was used to generate a  
CC cloning cassette, which was used to clone the sequence blocks of the  
CC present invention.  
XX  
SQ Sequence 62 AA;  
  
Query Match 31.2%; Score 5; DB 21; Length 62;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 9 FIIFW 13  
Db 53 FIIFW 57  
  
RESULT 12  
AAB84314  
ID AAB84314 standard; Protein; 62 AA.  
XX  
AC AAB84314;  
XX  
DT 22-AUG-2001 (first entry)  
XX  
DE Amino acid sequence of a human CD28 protein fragment.  
XX  
KW Stimulatory primary signalling motif; immune cell; signal transduction;  
KW chimeric receptor; inflammatory disease; autoimmune disease; asthma;  
KW eczema; congenital disease; cystic fibrosis; sickle cell anemia;  
KW dermatological disease; psoriasis; neurological disease;  
KW multiple sclerosis; transplant-related disease; metabolic disease;  
KW organ transplant rejection; graft versus host disease;  
KW idiopathic disease; diabetes; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200132709-A2.  
XX  
PD 10-MAY-2001.  
XX  
PF 01-NOV-2000; 2000WO-GB04183.  
XX  
PR 01-NOV-1999; 99GB-0025848.  
XX  
PA (CELL-) CELLTech CHIROSCEINCE LTD.

XX PI Finney HM, Lawson ADG;  
 XX WPI; 2001-389718/41.  
 DR N-PSDB; AAH24845.  
 XX  
 PT Novel cytoplasmic signalling protein and chimeric receptor protein,  
 PT useful for treating HIV infection, asthma, eczema, psoriasis, multiple  
 PT sclerosis, contain non-natural stimulatory primary signalling motif -  
 XX  
 PS Example 1; Fig 2; 45pp; English.  
 XX  
 CC The present sequence represents a human CD28 extracellular spacer and  
 CC transmembrane region. Secondary signalling sequences derived from CD28  
 CC can be linked to a non-natural stimulatory primary signalling motif to  
 CC produce a cytoplasmic signalling sequence. The primary motif is efficient  
 CC at mediating immune cell signal transduction, particularly when  
 CC incorporated in an intracellular signalling domain of a chimeric  
 CC receptor. The primary signalling motif can be combined in any way so as  
 CC to achieve the desired level of activation (or inhibition) of a number  
 CC of secondary messenger cascades. The signalling motifs are useful in  
 CC therapy and in the manufacture of medicament for treating or preventing  
 CC disease in humans or animals. They are useful for treating human  
 CC patients suffering from infectious diseases e.g. human immunodeficiency  
 CC virus (HIV) infections, inflammatory/autoimmune diseases such as asthma  
 CC and eczema, congenital diseases e.g. cystic fibrosis, sickle cell anemia,  
 CC dermatological diseases e.g. psoriasis, neurological diseases e.g.  
 CC multiple sclerosis, transplant-related disease e.g. organ transplant  
 CC rejection, graft versus host disease, metabolic/idiopathic disease  
 CC e.g. diabetes, and cancer.  
 XX  
 SQ Sequence 62 AA;

Query Match 31.2%; Score 5; DB 22; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
 |||||  
 Db 53 FIIFW 57

RESULT 13  
 AAB98753  
 ID AAB98753 standard; Protein; 62 AA.

XX AAB98753;  
 AC  
 XX  
 DT 07-AUG-2001 (first entry)  
 XX  
 DE Human CD28 extracellular spacer and transmembrane region.  
 XX  
 KW Human; CD28; primary signalling motif; T cell receptor; TCR;  
 KW sequence block; SB; immunosuppressive; secondary signalling sequence;  
 KW antimicrobial; anti-inflammatory; dermatological; neuroprotective;  
 KW cytostatic; anti-HIV; antiasthmatic; antiskinning; antipsoriatic;  
 KW antidiabetic; gene therapy; diabetes; immune cell signal transduction;  
 KW infection; inflammation; cancer; autoimmune disease; congenital disease;  
 KW psoriasis; neurological disease; organ transplant rejection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200132867-A1.  
 XX  
 PD 10-MAY-2001.  
 XX  
 PF 01-NOV-2000; 2000WO-GB04193.  
 XX  
 PR 01-NOV-1999; 99GB-0025853.  
 XX  
 PA (CELL-) CELLTECH CHIROSCEINCE LTD.  
 XX  
 PI Finney HM, Lawson ADG;

XX  
 DR WPI; 2001-328791/34.  
 XX

PT New nucleic acids encoding polypeptides with expanded primary signalling  
 PT motifs, for use in gene therapy, particularly for treating or  
 PT preventing infections, inflammations or autoimmune diseases in humans  
 PT -

XX Example 1; Fig 2; 43pp; English.

XX The invention relates to novel primary signalling motifs containing  
 CC a consensus amino acid sequence. These motifs are extremely  
 CC efficient at mediating immune cell signal transduction, particularly  
 CC when incorporated into an intracellular signalling domain of a chimeric  
 CC receptor. Nucleic acids that encode, and polypeptides that contain,  
 CC these primary signalling motifs are useful in medicine and research.  
 CC They are useful in therapy, or in the manufacture of a medicament for  
 CC treating or preventing disease in humans or in animals. These diseases  
 CC include infections (e.g. HIV (human immunodeficiency virus) infection),  
 CC inflammatory or autoimmune diseases (e.g. asthma or eczema), congenital  
 CC diseases (e.g. cystic fibrosis or sickle cell anaemia), dermatological  
 CC diseases (e.g. psoriasis), neurological diseases (e.g. multiple  
 CC sclerosis), organ transplant rejection or graft-versus-host disease, or  
 CC metabolic/idiopathic diseases (e.g. diabetes or cancer). The  
 CC present sequence is the human CD28 extracellular spacer and  
 CC transmembrane region used in the construction of a cloning  
 CC cassette for generating sequence blocks of primary and secondary  
 CC signalling motifs. Primary signalling motifs are sequences that  
 CC transduce either a stimulatory or an inhibitory signal, which regulates  
 CC primary activation of the T cell receptor (TCR) complex. Secondary  
 CC motifs impart secondary or co-stimulatory signalling capacity to a  
 CC molecule in T cells.

XX Sequence 62 AA;

Query Match 31.2%; Score 5; DB 22; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
 |||||  
 Db 53 FIIFW 57

RESULT 14  
 AAB98800  
 ID AAB98800 standard; Protein; 62 AA.

XX AAB98800;  
 AC  
 XX  
 DT 08-AUG-2001 (first entry)  
 XX  
 DE Human CD28 extracellular spacer and transmembrane region.  
 XX  
 KW Human; anti-HIV; anti-inflammatory; antiasthmatic; dermatological;  
 KW antiskinning; antipsoriatic; neuroprotective; immunosuppressive;  
 KW antidiabetic; cytostatic; HIV infection; inflammation;  
 KW autoimmune disease; cystic fibrosis; sickle cell anaemia; psoriasis;  
 KW neurological disease; organ transplant rejection; diabetes; cancer;  
 KW graft-versus-host disease; adaptor receptor protein; CD28;  
 KW extracellular spacer; transmembrane region.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200132866-A2.  
 XX  
 PD 10-MAY-2001.  
 XX  
 PF 01-NOV-2000; 2000WO-GB04189.  
 XX  
 PR 01-NOV-1999; 99GB-0025854.  
 XX  
 PA (CELL-) CELLTECH CHIROSCEINCE LTD.

XX PI Finney HM, Lawson ADG;  
XX XX WPI; 2001-328790/34.  
DR XX  
XX  
PT Novel polynucleotide encoding adaptor receptor protein useful for  
PT treating human immunodeficiency virus (HIV) infection, asthma, cystic  
PT fibrosis, multiple sclerosis, organ transplant rejection, diabetes and  
PT cancer.  
XX  
PS Example 1; Fig 2; 52pp; English.  
XX  
CC The invention relates to a novel nucleic acid encoding an adaptor  
CC receptor protein comprising an extracellular ligand-binding domain, a  
CC transmembrane domain and an intracellular signalling domain. The  
CC intracellular signalling domain comprises the cytoplasmic portion of at  
CC least one adaptor protein, and the extracellular ligand-binding domain  
CC is not CD8 or a major histocompatibility complex (MHC) class I protein.  
CC The adaptor receptor protein and the nucleic acid encoding it are useful  
CC in therapy. They are useful in the manufacture of a medicament for the  
CC treatment or prevention of disease in humans and animals. They are useful  
CC in the treatment of infectious diseases (e.g. HIV infection),  
CC inflammatory and autoimmune diseases (e.g. asthma and eczema),  
CC congenital diseases (e.g. cystic fibrosis and sickle cell anaemia),  
CC dermatological diseases (e.g. psoriasis), neurological diseases (e.g.)  
CC multiple sclerosis), organ transplant rejection, graft-versus-host  
CC disease and metabolic/idiopathic diseases such as diabetes and cancer.  
CC The present sequence is the human CD28 extracellular spacer  
CC and transmembrane region. It was used in the construction of a cloning  
CC cassette for cloning adaptor molecules of the invention.  
XX  
SQ Sequence 62 AA;

Query Match 31.2%; Score 5; DB 22; Length 62;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY -9 FIIIFW 13  
|||||  
Db 53 FIIIFW 57

## RESULT 15

ID AAB28678 standard; Protein; 62 AA.  
XX  
AC AAB28678;

DT 14-FEB-2001 (first entry)  
XX  
DE Protein encoded by cloning cassette system PCR fragment.

DE Membrane-associated protein; antiviral; antibacterial; antiparasitic;  
KW immunomodulatory; anticancer; antiinflammatory; antiasthmatic;  
KW antidiabetic; neuroprotective; chimeric receptor; infection;  
KW inflammatory disorder; autoimmune disorder; cancer; allergy; asthma;  
KW eczema; cystic fibrosis; sickle cell anaemia; psoriasis;  
KW multiple sclerosis; organ transplant rejection; diabetes.

OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO200063373-A1.  
XX  
PD 26-OCT-2000.  
XX  
PF 17-APR-2000; 2000WO-GB01471.  
XX  
PR 16-APR-1999; 99GB-0008816.  
XX  
PA (CLLT ) CELLTech THERAPEUTICS LTD.  
XX  
PI Finney HM, Lawson ADG;

XX  
DR WPI; 2001-015774/02.  
DR N-PSDB; AAC65397.

XX  
PT Altering the properties or level of expression of membrane-associated  
PT proteins, e.g., to change responses to cell surface antigens or the  
PT sensitivity of intracellular signaling -

PS Example 1; Fig 2; 47pp; English.

XX  
CC The nucleotide sequence encoding the present protein was used in the  
CC construction of chimeric receptors with different binding, extracellular  
CC spacer, transmembrane and signalling components. The properties and level  
CC of expression of a membrane-associated protein may be altered by  
CC substituting a transmembrane region or a membrane-anchoring region for  
CC transmembrane or membrane-anchoring regions that are not naturally part  
CC of the protein. The relative response of membrane-associated proteins to  
CC cell surface-associated antigen versus antigen in solution, and the  
CC sensitivity of intracellular signaling mediated by membrane-associated  
CC proteins can be altered. They can be important in treatment of, e.g. HIV  
CC infection, bacterial infections, parasitic infections,  
CC inflammatory/autoimmune disorders (e.g. rheumatoid arthritis,  
CC osteoarthritis or inflammatory bowel disease), cancer, allergic/atopic  
CC diseases (e.g. asthma or eczema), congenital disorders (e.g. cystic  
CC fibrosis or sickle cell anaemia), dermatological disorders (e.g.  
CC psoriasis), neurological disorders (e.g. multiple sclerosis), organ  
CC transplant rejection, graft-versus-host diseases, or metabolic/idiopathic  
CC diseases (e.g. diabetes). The polynucleotide encoding the present  
CC sequence starts with a SpeI site and consists of the extracellular spacer  
CC h.CD28, the human CD28 transmembrane region, a stop codon, and finishes  
CC with an EcoRI site.

SQ Sequence 62 AA;

Query Match 31.2%; Score 5; DB 22; Length 62;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIIFW 13  
|||||  
Db 53 FIIIFW 57

## RESULT 16

AAU98026  
ID AAU98026 standard; Protein; 62 AA.

XX  
AC AAU98026;

DT 27-AUG-2002 (first entry)  
XX

DE Human IgG1 hinge/CD28 transmembrane domain.

XX  
KW Human; IgG1 hinge; CD28; transmembrane domain; CD137; TNFR;  
KW tumour necrosis factor receptor; cytoplasmic signalling molecule;  
KW gene therapy; chimeric receptor; human immunodeficiency virus infection;  
KW HIV; inflammatory disease; autoimmune disease; asthma; eczema; cancer;  
KW congenital disease; cystic fibrosis; sickle cell anaemia; diabetes;  
KW dermatological disease; psoriasis; neurological disease;  
KW multiple sclerosis; transplantation associated disease;  
KW organ transplant rejection; graft-versus-host disease; metabolic disease;  
KW idiopathic disease.

XX Homo sapiens.  
OS Synthetic.

XX WO200233101-A1.

PN 25-APR-2002.

XX 16-OCT-2001; 2001WO-GB04611.

XX 16-OCT-2000; 2000GB-0025307.



XX (CELL-) CELLTech R & D LTD.  
XX PI Finney HM, Lawson ADG;  
XX WPI; 2002-463315/49.  
DR N-PSDB; ABK52921.  
DR  
XX Novel nucleic acid encoding cytoplasmic signalling molecule comprising  
PT at least one cytoplasmic signalling sequence derived from CD137, useful  
PT for mediating signalling when employed as part of chimaeric receptor  
PT protein  
XX  
PS Example 1; Fig 2; 39pp; English.  
XX  
CC The invention relates to a nucleic acid (I) encoding a cytoplasmic  
CC signalling molecule comprising at least two cytoplasmic signalling  
CC sequences where at least one cytoplasmic signalling sequence derived from  
CC CD137 (a member of the tumour necrosis receptor family, TNFR).  
CC Also included are a nucleic acid (II) encoding a chimaeric receptor  
CC protein which comprises an extracellular ligand-binding domain, a  
CC transmembrane domain and a cytoplasmic signalling domain, where the  
CC cytoplasmic signalling domain is encoded by nucleic acid (I), a vector  
CC comprising a nucleic acid (I) or (II), a host cell containing nucleic  
CC acid (I) or (II) or the vector, a peptide or polypeptide comprising a  
CC cytoplasmic signalling molecule encoded by nucleic acid (I) and a  
CC chimaeric receptor protein encoded by nucleic acid (II). The novel  
CC cytoplasmic signalling molecules encoded by nucleic acid can be used  
CC either by themselves or, as a component part of a larger protein such as  
CC a chimaeric receptor. The cytoplasmic signalling molecules are used  
CC preferentially to mediate signalling when employed as a cytoplasmic  
CC signalling domain of a chimaeric receptor protein. The cytoplasmic  
CC signalling molecules, chimaeric receptors and nucleic acids encoding  
CC them are preferably useful for treating human immunodeficiency virus  
CC (HIV) infection, inflammatory disease e.g. cystic fibrosis, or sickle cell anaemia,  
CC eczema, congenital disease e.g. psoriasis, neurological disease e.g. multiple  
CC dermatological disease e.g. psoriasis, organ transplant  
CC sclerosis, transplantation associated diseases e.g. organ transplant  
CC rejection, or graft-versus-host disease, metabolic/idiopathic disease,  
CC e.g. diabetes, or cancer. The present sequence represents a chimaeric  
CC protein consisting of the hinge region of human IgG1 and a region of  
CC human CD28 containing a transmembrane domain. This region is used in the  
CC chimaeric receptor constructs of the invention.  
XX  
SQ Sequence 62 AA;  
  
Query Match 31.2%; Score 5; DB 23; Length 62;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 9 FIIWF 13  
Db 53 FIIWF 57  
  
RESULT 17  
ID AAM92938 standard; Protein; 67 AA.  
XX  
AC AAM92938;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human digestive system antigen SEQ ID NO: 2287..  
XX  
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum.  
XX  
OS Homo sapiens.  
XX  
PN WO200155314-A2.  
XX

PD 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01324.  
PF 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PR (HUMA-) HUMAN GENOME SCI INC.  
PR Rosen CA, Barash SC, Ruben SM;  
PR WPI; 2001-502630/55.

DR N-PSDB; AAK88711.  
XX Polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases -  
XX Claim 11; SEQ ID NO 2287; 986pp; English.  
XX The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a digestive system antigen of  
XX the invention.  
SQ Sequence 67 AA;  
Query Match 31.2%; Score 5; DB 22; Length 67;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 DFIF 12  
Db 32 DFIF 36  
RESULT 18  
AAU20034  
ID AAU20034 standard; Protein; 67 AA.  
XX AC AAU20034;  
XX 06-DEC-2001 (first entry)  
DE Human liver associated polypeptide #65.  
XX Liver associated protein; human; mouse; rabbit; goat; horse; cat; dog;  
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic;  
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;  
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;  
KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;  
KW hyperproliferative disorder; breast; liver; cardiovascular disorder;  
KW cerebrovascular disorder; nervous system disorder; bacterial infection;  
KW fungal infection; viral infection; ocular disorder; endocrine disorder;  
KW gastrointestinal disorder; renal disorder; respiratory disorder;  
KW wound healing; skin aging; organ transplantation; tissue regeneration;  
XX anti-infertility.  
OS Homo sapiens.  
XX WO200155355-A1.  
PN 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01351.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-457728/49.  
N-PSDB; AAS31745.

Isolated nucleic acid molecule encoding a human liver related protein is used in preventing, treating or ameliorating disorders of the liver particularly cancer of the liver -

Claim 11; SEQ ID No 221; 526pp; English.

Sequences AAU19970-AAU20115 represent the liver associated polypeptides of the invention. Liver associated polypeptides and their associated polynucleotides are useful in the diagnosis, treatment and prevention of various types of disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A pathological condition can be determined by detecting the presence or absence of a mutation in a liver associated polynucleotide. The treatable disorders include autoimmune diseases such as rheumatoid arthritis, hyperproliferative disorders such as neoplasms of the breast or liver, cardiovascular disorders such as cardiac arrest, cerebrovascular disorders such as cerebral ischaemia, nervous system disorders such as Alzheimer's disease, infections caused by bacteria, viruses and fungi, ocular disorders such as corneal infection, endocrine

CC disorders such as premature labour and infertility, gastrointestinal  
CC disorders such as Crohn's disease, renal disorders such as  
CC glomerulonephritis and respiratory disorders such as asthma and pleurisy.  
CC The polypeptides can also be used to aid wound healing, to prevent skin  
CC aging due to sunburn, to maintain organs before transplantation, to  
CC regenerate tissues and in chemotaxis.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

Query Match 31.2%; Score 5; DB 22; Length 67;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 DFIIF 12  
Db 32 DFIIF 36

RESULT 19  
ABP40895  
ID ABP40895 standard; Protein; 67 AA.  
XX  
AC ABP40895;  
XX

DT 24-JUL-2002 (first entry)  
XX

DE Human liver antigen HLICL82, SEQ ID NO:221.  
XX

KW Human; liver antigen; liver disorder; hepatic disorder; infection;  
KW hepatitis; viral; parasitic; bacterial; fungal; inflammatory condition;  
KW cirrhosis; granulomatous hepatitis; toxin damage; drug damage;  
KW autoimmune disease; Wilson's disease; primary biliary cirrhosis;  
KW neoplastic disorder; cancer; tumour; portal hypertension;  
KW gastrointestinal disorder; hepatitis; drug screening; gene therapy;  
KW chromosome mapping; forensic analysis; antibody preparation;  
KW hepatotropic; cytostatic; antiinflammatory; virucide; antibacterial;  
KW fungicide; parasiticide; antidote; immunosuppressive.  
XX

OS Homo sapiens.  
XX

XX US2002042096-A1.  
XX

PD 11-APR-2002.  
XX

PF 17-JAN-2001; 2001US-0764887.  
XX

PR 31-JAN-2000; 2000US-179065P.  
PR 04-FEB-2000; 2000US-180628P.  
PR 28-JUN-2000; 2000US-214886P.  
PR 07-JUL-2000; 2000US-216647P.  
PR 07-JUL-2000; 2000US-216880P.  
PR 11-JUL-2000; 2000US-217487P.  
PR 11-JUL-2000; 2000US-217496P.  
PR 14-JUL-2000; 2000US-218290P.  
PR 26-JUL-2000; 2000US-220963P.  
PR 26-JUL-2000; 2000US-220964P.  
PR 14-AUG-2000; 2000US-224518P.  
PR 14-AUG-2000; 2000US-224519P.  
PR 14-AUG-2000; 2000US-225267P.  
PR 14-AUG-2000; 2000US-225268P.  
PR 14-AUG-2000; 2000US-225270P.  
PR 14-AUG-2000; 2000US-225447P.  
PR 14-AUG-2000; 2000US-225757P.  
PR 14-AUG-2000; 2000US-225758P.  
PR 22-AUG-2000; 2000US-226868P.  
PR 30-AUG-2000; 2000US-228924P.  
PR 01-SEP-2000; 2000US-229287P.  
PR 01-SEP-2000; 2000US-229343P.  
PR 01-SEP-2000; 2000US-229344P.  
PR 01-SEP-2000; 2000US-229345P.  
PR 05-SEP-2000; 2000US-229509P.

PR 05-SEP-2000; 2000US-229513P.  
PR 08-SEP-2000; 2000US-231413P.  
PR 21-SEP-2000; 2000US-234223P.  
PR 21-SEP-2000; 2000US-234274P.  
PR 25-SEP-2000; 2000US-234997P.  
PR 27-SEP-2000; 2000US-235834P.  
PR 29-SEP-2000; 2000US-236327P.  
PR 29-SEP-2000; 2000US-236367P.  
PR 29-SEP-2000; 2000US-236368P.  
PR 29-SEP-2000; 2000US-236369P.  
PR 02-OCT-2000; 2000US-236370P.  
PR 02-OCT-2000; 2000US-236802P.  
PR 02-OCT-2000; 2000US-237037P.  
PR 02-OCT-2000; 2000US-237038P.  
PR 02-OCT-2000; 2000US-237039P.  
PR 13-OCT-2000; 2000US-237040P.  
PR 20-OCT-2000; 2000US-239935P.  
PR 20-OCT-2000; 2000US-240960P.  
PR 20-OCT-2000; 2000US-241785P.  
PR 20-OCT-2000; 2000US-241809P.  
PR 01-NOV-2000; 2000US-244617P.  
PR 17-NOV-2000; 2000US-249299P.  
PR 08-DEC-2000; 2000US-251856P.  
PR 08-DEC-2000; 2000US-251868P.  
PR 08-DEC-2000; 2000US-251869P.  
XX

PA (ROSE/) ROSEN C A.  
PA (RUBE/) RUBEN S M.  
PA (BARA/) BARASH S C.  
XX

PI Rosen CA, Ruben SM, Barash SC;  
XX

XX WPI; 2002-381944/41.  
DR N-PSDB; ABN90100.  
XX

PT New nucleic acid encoding human liver antigens, useful for diagnosis,  
PT treatment and prevention of e.g. hepatitis and hepatic cancer, also  
PT related polypeptides and antibodies  
XX

PS Claim 11; SEQ ID No 221; 181pp; English.  
XX

CC The invention relates to 145 novel human liver antigens (ABP40831-  
CC ABP40975) and to cDNAs encoding them (ABN90036-ABN90180), and also  
CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
CC to the sequences of the invention. The invention additionally relates to  
CC recombinant vectors and host cells comprising human liver antigen  
CC polynucleotides, antibodies against human liver antigens, and the use of  
CC liver antigen polynucleotides and polypeptides in diagnosing, and the use of  
CC prognosing or preventing various disorders of the liver. Such conditions  
CC include viral infections (e.g., cytomegalovirus, Epstein-Barr virus,  
CC hepatitis A virus, hepatitis B virus and hepatitis C virus), parasitic  
CC infections (e.g., Clonorchis sinensis, Echinococcus granulosus and  
CC Entamoeba histolytica), and also bacterial and fungal infections. Other  
CC disorders that may be treated include inflammatory conditions (e.g.,  
CC cirrhosis and granulomatous hepatitis), damage caused by drugs or toxins,  
CC autoimmune diseases (e.g. Wilson's disease, primary biliary cirrhosis),  
CC neoplastic disorders (e.g., adenomas, haemangiomas and hepatocellular  
CC carcinoma), portal hypertension, or gastrointestinal disorders (e.g.,  
CC peptic ulcers, gastritis and peritoneal diseases). Liver antigen  
CC polypeptides and polynucleotides may also be used in screening for  
CC compounds which modulate liver antigen expression or activity. The  
CC polynucleotides may further be used for gene therapy, chromosome  
CC mapping, in the identification of individuals and in forensic analysis,  
CC and the polypeptides may be used as molecular weight markers or to  
CC prepare antibodies useful in disease diagnosis, drug targeting and  
CC phenotyping. The present sequence represents a human liver antigen of  
CC the invention.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO at seqdata.uspto.gov/sequence/  
XX

SQ Sequence 67 AA;



Query Match 31.2%; Score 5; DB 23; Length 67;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12  
| | | | |  
Db 32 DFIIIF 36

RESULT 20  
AAB58985  
ID AAB58985 standard; Protein; 68 AA.  
XX  
AC AAB58985;  
XX  
DT 27-MAR-2001 (first entry)  
XX  
DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 693.  
XX  
KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
KW nontropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
KW antidiabetic; antiinflammatory; antiulcer; vulnery; anticonvulsant;  
KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;  
KW Addison's disease; allergy; autoimmune haemolytic anaemia;  
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
KW cardiovascular disorder; wound healing; neurological disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200055173-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 08-MAR-2000; 2000WO-US05881.  
XX  
PR 12-MAR-1999; 99US-0124270.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM;  
XX  
DR WPI; 2000-611515/58.  
DR N-PSDB; AAF21888.  
XX  
PT New human breast and ovarian cancer associated gene sequences and the  
PT polypeptides encoded by these genes, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
PT disorders and neurological diseases -  
XX  
PS Claim 11; Page 1147; 1299pp; English.  
XX  
CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
CC associated with breast and ovarian cancer. Included in the invention are  
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
CC isolation and characterisation of the DNA and protein sequences of the  
CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
CC or antagonist sequences exhibit cytostatic; immunosuppressive;  
CC nontropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
CC antidiabetic; antiinflammatory; antiulcer; vulnery; anticonvulsant;  
CC antibacterial; antifungal; antiparasitic and cardiant activity. The  
CC polynucleotide and protein sequences are used in the diagnosis of cancer,  
CC particularly breast and ovarian cancer. The nucleic acid sequences,  
CC proteins, agonists and antagonists may also be used in the diagnosis,  
CC prevention and treatment of immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; cardiovascular disorders such as  
CC myocardial ischaemias; wound healing; neurological diseases such as  
CC cerebral anoxia and epilepsy; and infectious diseases.  
XX  
SQ Sequence 68 AA;

Query Match 31.2%; Score 5; DB 21; Length 68;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12  
| | | | |  
Db 47 DFIIIF 51

RESULT 21  
AAR94651  
ID AAR94651 standard; Protein; 69 AA.  
XX  
AC AAR94651;  
XX  
DT 08-DEC-1996 (first entry)  
XX  
DE Mouse Fas-associated protein PTP-BAS type 5b (MFAP23).  
XX  
KW Fas-associated protein; tumour necrosis factor receptor; PTP-BAS;  
KW apoptosis; FAP; cell surface protein; autoimmune disease; HIV virus;  
KW hybridoma; cell death.  
XX  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 34..69  
FT /note= "AAS diverging from PTP-BAS family"  
XX  
PN WO9534661-A1.  
XX  
PD 21-DEC-1995.  
XX  
PF 14-JUN-1995; 95WO-US07583.  
XX  
PR 27-MAR-1995; 95US-0410804.  
PR 14-JUN-1994; 94US-0259514.  
XX  
PA (LJOL-) LA JOLLA CANCER RES FOUND.  
XX  
PI Reed JC, Sato T;  
XX  
DR WPI; 1996-049689/05.  
DR N-PSDB; AAT18385.  
XX  
PT New Fas associated proteins PTP-BAS types 4 and 5 - involved in  
PT programmed cell death, used for modulating apoptosis, e.g. for  
PT treating cancer, and for identifying other modulators  
XX  
PS Claim 17; Fig.16; 84pp; English.  
XX  
CC Regulation of apoptosis can be used in the treatment of cancer,  
CC autoimmune disease and viral disease (e.g. HIV virus infections),  
CC also to prolong survival of e.g. hybridoma cells in culture  
CC resulting in an increase in bioproduct yield. This protein may  
CC also be used in a method for diagnosing diseases associated with  
CC increased or decreased levels of Fas-associated protein.  
XX  
SQ Sequence 69 AA;  
Query Match 31.2%; Score 5; DB 17; Length 69;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
| | | | |  
Db 58 IDFII 62

RESULT 22  
AAG00588  
ID AAG00588 standard; Protein; 72 AA.

```
XX AC AAG00588;
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein, SEQ ID NO: 4669.
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping.
XX OS Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-0200610.
XX PR 26-FEB-1999; 99US-0122487.
XX PA (GEST ) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI; 2000-500381/45.
XX DR N-PSDB; AAC00594.
XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX PS Claim 13; SEQ ID 4669; 71pp + CD-ROM; English.
XX CC The present sequence is a polypeptide encoded by one of a large number
XX CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
XX CC were prepared from total human RNAs or polyA+ RNAs derived from 30
XX CC different tissues. EST sequences usually correspond mainly to the 3'
XX CC untranslated region (UTR) of the mRNA because they are often obtained
XX CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
XX CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
XX CC those cases where longer cDNA sequences have been obtained, the full 5'
XX CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
XX CC ends and can therefore be used to obtain full length cDNAs and genomic
XX CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
XX CC chromosome mapping procedures. They are used to obtain upstream
XX CC regulatory sequences and to design expression and secretion vectors.
XX SQ Sequence 72 AA;
Query Match 31.2%; Score 5; DB 21; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 IDFII 11
Db 12 IDFII 16
RESULT 23
ABP40821
ID ABP40821 standard; Protein; 74 AA.
XX AC ABP40821;
XX DT 24-JUL-2002 (first entry)
XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5666.
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX KW antibacterial; gene therapy.
XX OS Staphylococcus epidermidis.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
```

```
PN US6380370-B1.
XX 30-APR-2002.
XX 13-AUG-1998; 98US-0134001.
XX 14-AUG-1997; 97US-055779P.
XX 08-NOV-1997; 97US-064964P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Doucette-Stamm LA, Bush D;
XX WPI; 2002-381255/41.
XX N-PSDB; ABN93366.
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX polypeptide, useful for diagnosing and treating bacterial infections -
XX PS Disclosure; SEQ ID 5666; 267pp; English.
XX CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX CC antibacterial activity and can be used in gene therapy. The sequences
XX CC can also be used in the diagnosis and treatment of bacterial infections,
XX CC particularly S. epidermidis infections. The sequences can be used to
XX CC screen for compounds able to interfere with the S. epidermidis life
XX CC cycle or inhibit S. epidermidis infection.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC USPTO web site.
XX SQ Sequence 74 AA;
Query Match 31.2%; Score 5; DB 23; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 FIIIFW 13
Db 40 FIIIFW 44
RESULT 24
ABG27612
ID ABG27612 standard; Protein; 76 AA.
XX AC ABG27612;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #27603.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
```

XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;  
PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;  
XX WPI; 1997-424969/39.  
DR N-PSDB; AAT84086.  
XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used  
PT to isolate antimicrobial compounds, and in vaccines against S.  
PT aureus infection  
XX Claim 6; Page 504; 989pp; English.  
PS The present sequence represents a Staphylococcus aureus protein, that,  
XX based on homology with an Escherichia coli protein, is believed to be a  
CC large conductance mechanosensitive channel. The DNA sequence was isolated  
CC from a library of clones of S. aureus WCUH 29 in Escherichia coli. The  
CC DNA sequences can be used in the construction of ribozymes and antisense  
CC sequences to control the expression of Staphylococcal genes. The DNA  
CC sequence is also useful as a source of regulatory elements for the  
CC control of bacterial gene expression. The present protein may be used  
CC to produce vaccines to enable a host to produce specific antibodies  
CC with antibacterial action. These vaccines and antibodies would protect  
CC a host against invasion by S. aureus, and conditions relating to  
CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled  
CC skin syndrome, and toxic shock syndrome.  
XX Sequence 81 AA;  
SQ Query Match 31.2%; Score 5; DB 18; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11  
Db 32 IDFI 36

RESULT 26  
AAW28161  
ID AAW28161 standard; Protein; 81 AA.  
XX AC AAW28161;  
XX DT 27-AUG-1998 (first entry)  
XX DE Amino acid sequence of a large conductance mechanosensitive channel.  
XX KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;  
KW Staphylococcal gene; regulatory element; bacterial gene expression;  
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;  
KW toxic shock syndrome.  
XX OS Staphylococcus aureus.  
XX FH Key Location/Qualifiers  
FT Misc-difference 1.81  
FT /note= "residues designated x are not defined in  
FT the specification"  
XX PN WO9730070-A1.  
XX PD 21-AUG-1997.  
XX PF 19-FEB-1997; 97WO-US02318.  
XX PR 20-FEB-1996; 96US-0011888.  
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.

DR N-PSDB; AAS91799.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 20; SEQ ID No 57971; 103pp; English.  
PS The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 76 AA;  
SQ Query Match 31.2%; Score 5; DB 22; Length 76;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 9 GIDFI 13

RESULT 25  
AAW28161  
ID AAW28161 standard; Protein; 81 AA.  
XX AC AAW28161;  
XX DT 27-AUG-1998 (first entry)  
XX DE Amino acid sequence of a large conductance mechanosensitive channel.  
XX KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;  
KW Staphylococcal gene; regulatory element; bacterial gene expression;  
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;  
KW toxic shock syndrome.  
XX OS Staphylococcus aureus.  
XX FH Key Location/Qualifiers  
FT Misc-difference 1.81  
FT /note= "residues designated x are not defined in  
FT the specification"  
XX PN WO9730070-A1.  
XX PD 21-AUG-1997.  
XX PF 19-FEB-1997; 97WO-US02318.  
XX PR 20-FEB-1996; 96US-0011888.  
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.

PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249266.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.  
N-PSDB; AAL00559.

Isolated nucleic acid molecule encoding a reproductive system antigen -  
is used in preventing, treating or ameliorating a medical condition -

Claim 11; SEQ ID NO 3247; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a  
number of human reproductive system related antigens. These can be used  
in the prevention and treatment of reproductive system disorders,  
including cancer. The present sequence is a protein of the invention.

Sequence 82 AA;

Query Match

Best Local Similarity 31.2%; Score 5; DB 22; Length 82;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





DT	07-NOV-2001	(first entry)	
XX	Human	immune/haematopoietic antigen	SEQ ID NO:18336.
DE	Human	immune; haematopoietic; immune/haematopoietic antigen; cancer;	
XX	cytostatic; gene therapy; vaccine; metastasis.		
KW	Homo sapiens.		
KW	WO200157182-A2.		
XX	09-AUG-2001.		
XX	17-JAN-2001;	2001WO-US01354.	
PF	31-JAN-2000;	2000US-0179065.	
XX	04-FEB-2000;	2000US-0180628.	
PR	24-FEB-2000;	2000US-0184664.	
PR	02-MAR-2000;	2000US-0186350.	
PR	16-MAR-2000;	2000US-0189874.	
PR	17-MAR-2000;	2000US-0190076.	
PR	18-APR-2000;	2000US-0198123.	
PR	19-MAY-2000;	2000US-0205515.	
PR	07-JUN-2000;	2000US-0209467.	
PR	28-JUN-2000;	2000US-0214886.	
PR	30-JUN-2000;	2000US-0215135.	
PR	07-JUL-2000;	2000US-0216647.	
PR	07-JUL-2000;	2000US-0216880.	
PR	11-JUL-2000;	2000US-0217487.	
PR	11-JUL-2000;	2000US-0217496.	
PR	14-JUL-2000;	2000US-0218290.	
PR	26-JUL-2000;	2000US-0220963.	
PR	26-JUL-2000;	2000US-0220964.	
PR	14-AUG-2000;	2000US-0224518.	
PR	14-AUG-2000;	2000US-0224519.	
PR	14-AUG-2000;	2000US-0225213.	
PR	14-AUG-2000;	2000US-0225214.	
PR	14-AUG-2000;	2000US-0225266.	
PR	14-AUG-2000;	2000US-0225267.	
PR	14-AUG-2000;	2000US-0225268.	
PR	14-AUG-2000;	2000US-0225270.	
PR	14-AUG-2000;	2000US-0225447.	
PR	14-AUG-2000;	2000US-0225757.	
PR	14-AUG-2000;	2000US-0225758.	
PR	14-AUG-2000;	2000US-0225759.	
PR	18-AUG-2000;	2000US-0226279.	
PR	22-AUG-2000;	2000US-0226681.	
PR	22-AUG-2000;	2000US-0226868.	
PR	23-AUG-2000;	2000US-0227182.	
PR	30-AUG-2000;	2000US-0227009.	
PR	01-SEP-2000;	2000US-0228924.	
PR	01-SEP-2000;	2000US-0229287.	
PR	01-SEP-2000;	2000US-0229343.	
PR	01-SEP-2000;	2000US-0229344.	
PR	05-SEP-2000;	2000US-0229509.	
PR	05-SEP-2000;	2000US-0229513.	
PR	06-SEP-2000;	2000US-0230437.	
PR	06-SEP-2000;	2000US-0230438.	
PR	08-SEP-2000;	2000US-0231242.	
PR	08-SEP-2000;	2000US-0231243.	
PR	08-SEP-2000;	2000US-0231244.	
PR	08-SEP-2000;	2000US-0231413.	
PR	08-SEP-2000;	2000US-0231414.	
PR	08-SEP-2000;	2000US-0232080.	
PR	12-SEP-2000;	2000US-0232081.	
PR	14-SEP-2000;	2000US-0231968.	
PR	14-SEP-2000;	2000US-0232397.	
PR	14-SEP-2000;	2000US-0232398.	
PR	14-SEP-2000;	2000US-0232399.	
PR	14-SEP-2000;	2000US-0232400.	
PR	14-SEP-2000;	2000US-0232401.	
PR	14-SEP-2000;	2000US-0233063.	
PR	14-SEP-2000;	2000US-0233064.	
PR	14-SEP-2000;	2000US-0233065.	
PR	21-SEP-2000;	2000US-0234223.	
PR	21-SEP-2000;	2000US-0234274.	
PR	25-SEP-2000;	2000US-0234997.	
PR	25-SEP-2000;	2000US-0234998.	
PR	26-SEP-2000;	2000US-0235484.	
PR	27-SEP-2000;	2000US-0235834.	
PR	27-SEP-2000;	2000US-0235836.	
PR	29-SEP-2000;	2000US-0236327.	
PR	29-SEP-2000;	2000US-0236367.	
PR	29-SEP-2000;	2000US-0236368.	
PR	29-SEP-2000;	2000US-0236369.	
PR	29-SEP-2000;	2000US-0236370.	
PR	02-OCT-2000;	2000US-0237037.	
PR	02-OCT-2000;	2000US-0237038.	
PR	02-OCT-2000;	2000US-0237039.	
PR	02-OCT-2000;	2000US-0237040.	
PR	13-OCT-2000;	2000US-0239935.	
PR	13-OCT-2000;	2000US-0239937.	
PR	20-OCT-2000;	2000US-0240960.	
PR	20-OCT-2000;	2000US-0241221.	
PR	20-OCT-2000;	2000US-0241785.	
PR	20-OCT-2000;	2000US-0241786.	
PR	20-OCT-2000;	2000US-0241787.	
PR	20-OCT-2000;	2000US-0241808.	
PR	20-OCT-2000;	2000US-0241809.	
PR	01-NOV-2000;	2000US-0241826.	
PR	08-NOV-2000;	2000US-0244617.	
PR	08-NOV-2000;	2000US-0246474.	
PR	08-NOV-2000;	2000US-0246475.	
PR	08-NOV-2000;	2000US-0246476.	
PR	08-NOV-2000;	2000US-0246477.	
PR	08-NOV-2000;	2000US-0246478.	
PR	08-NOV-2000;	2000US-0246523.	
PR	08-NOV-2000;	2000US-0246524.	
PR	08-NOV-2000;	2000US-0246525.	
PR	08-NOV-2000;	2000US-0246526.	
PR	08-NOV-2000;	2000US-0246527.	
PR	08-NOV-2000;	2000US-0246528.	
PR	08-NOV-2000;	2000US-0246532.	
PR	08-NOV-2000;	2000US-0246609.	
PR	08-NOV-2000;	2000US-0246610.	
PR	08-NOV-2000;	2000US-0246611.	
PR	08-NOV-2000;	2000US-0246613.	
PR	17-NOV-2000;	2000US-0249207.	
PR	17-NOV-2000;	2000US-0249208.	
PR	17-NOV-2000;	2000US-0249209.	
PR	17-NOV-2000;	2000US-0249210.	
PR	17-NOV-2000;	2000US-0249211.	
PR	17-NOV-2000;	2000US-0249212.	
PR	17-NOV-2000;	2000US-0249213.	
PR	17-NOV-2000;	2000US-0249214.	
PR	17-NOV-2000;	2000US-0249215.	
PR	17-NOV-2000;	2000US-0249216.	
PR	17-NOV-2000;	2000US-0249217.	
PR	17-NOV-2000;	2000US-0249218.	
PR	17-NOV-2000;	2000US-0249244.	
PR	17-NOV-2000;	2000US-0249245.	
PR	17-NOV-2000;	2000US-0249264.	
PR	17-NOV-2000;	2000US-0249265.	
PR	17-NOV-2000;	2000US-0249297.	
PR	17-NOV-2000;	2000US-0249299.	
PR	17-NOV-2000;	2000US-0249300.	
PR	01-DEC-2000;	2000US-0250160.	
PR	05-DEC-2000;	2000US-0250391.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	06-DEC-2000;	2000US-0256719.	
PR	08-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	

PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-483426/52.  
 XX N-PSDB; AAK63524.  
 DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 XX Claim 11; SEQ ID NO 18336; 3071pp + Sequence Listing; English.  
 PS AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 92 AA;  
 Query Match 31.2%; Score 5; DB 22; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 ANCGI 7  
 Db 18 ANCGI 22  
 RESULT 30  
 AAY24470  
 ID AAY24470 standard; Protein; 103 AA.  
 XX  
 AC AAY24470;  
 XX 24-SEP-1999 (first entry)  
 DT Human CD28 gene protein sequence #2.  
 XX  
 DE CD28; inhibition; antisense oligonucleotide; interleukin 2; IL-2;  
 XX immune system mediated disease; gamma-interferon; IL-8.  
 KW Homo sapiens.  
 XX US5932556-A.  
 PN 03-AUG-1999.  
 XX 18-SEP-1995; 95US-0529878.  
 PD 18-SEP-1995; 95US-0529878.  
 XX (TAMR/) TAM R C.  
 PA  
 XX

PI Tam RC;  
 XX WPI; 1999-443609/37.  
 DR N-PSDB; AAX90349.  
 XX Treatment of immune system-mediated diseases by inhibiting  
 PT expression of CD28, IL-2, gamma-interferon or IL-8  
 XX Disclosure; Fig 1C; 45pp; English.  
 PS The present invention describes a method for inhibiting the expression  
 XX of CD28, IL-2, gamma-interferon or IL-8 in a mammal. The method  
 CC comprises subcutaneous administration of an oligonucleotide (OGN).  
 CC AAX90288 to AAX90291 represent specifically claimed OGNs for use in the  
 CC method. The OGNs are used for the treatment of immune system-mediated  
 CC diseases. The present sequence represents a protein sequence of the  
 CC human CD28 gene from the present invention.  
 XX  
 SQ Sequence 103 AA;  
 Query Match 31.2%; Score 5; DB 20; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 FIIFW 13  
 Db 57 FIIFW 61  
 RESULT 31  
 AAM94419  
 ID AAM94419 standard; Protein; 104 AA.  
 XX  
 AC AAM94419;  
 XX 21-NOV-2001 (first entry)  
 DT Human reproductive system related antigen SEQ ID NO: 3077.  
 XX  
 DE Human; reproductive system related antigen; reproductive system disorder;  
 KW cancer; gene therapy.  
 KW Homo sapiens.  
 OS WO200155320-A2.  
 PN 02-AUG-2001.  
 XX 17-JAN-2001; 2001WO-US01339.  
 PF 31-JAN-2000; 2000US-0179065.  
 XX 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249254.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

N-PSDB; AAL00389.

Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition -

Claim 11; SEQ ID NO 3077; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a protein of the invention.

Sequence 104 AA;

Query Match 31.2%; Score 5; DB 22; Length 104;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
Db 31 IDFII 35

RESULT 32  
ABG30130  
ID ABG30130 standard; Protein; 111 AA.  
XX



AC ABG30130;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #30121.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS94317.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 20; SEQ ID No 60489; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 111 AA;  
  
Query Match 31.2%; Score 5; DB 22; Length 111;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 ANCGI 7  
Db 104 ANCGI 108  
  
RESULT 33  
AAG40079  
ID AAG40079 standard; Protein; 112 AA.  
XX  
AC AAG40079;  
XX

DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 49681.  
XX  
KW protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 21-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 31.2%; Score 5; DB 21; Length 112;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
Db 7 IDFII 11

RESULT 34  
ABB09650  
ID ABB09650 standard; Protein; 118 AA.  
XX  
AC ABB09650;

XX  
DT 29-MAY-2002 (first entry)

DE Amino acid sequence of gene encoded mutacin I (Muta) peptide Orfz.  
XX Lantibiotic; mutacin I; Muta; Streptococcus mutans CH43; probiotic;  
KW antimicrobial; multiple drug resistant bacteria; MutR; Muta; Muta';  
KW MutB; MutC; MutD; MutP; MutT; Orfz.

```

11-APR-2000; 2000FR-0004629.
(INSP ) INST PASTEUR.
Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
Daniels J, Gobel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
Rose M, Voss H;
WPI; 2002-010914/01.
Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
and prevention of Listeria and related bacterial infections, and
related polypeptides -
Claim 6; SEQ ID No 897; 192pp; French.
The present invention relates to the genome sequence of Listeria
monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
it are useful for selecting probes and primers for detecting genes in L.
monocytogenes and related organisms, and for studying genetic
polymorphisms and other genomes. The present sequence is a protein
encoded by the genome sequence of the present invention. Proteins
expressed from the genome sequence are useful for raising specific
antibodies, identification of L. monocytogenes and related organisms, and
for biosynthesis and biodegradation, especially biosynthesis of Vitamin
B12. The genome sequence and proteins encoded by it are also useful for
selecting compounds that regulate gene expression and cell replication
and modulate L. monocytogenes-related diseases. In addition, the genome
sequence and proteins encoded by it are useful in pharmaceutical and
vaccines compositions for the treatment or prevention of infections by L.
monocytogenes and related organisms.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 120 AA;
Query Match 31.2%; Score 5; DB 23; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 IDFII 11
Db 27 IDFII 31
RESULT 36
ABP40387
ID ABP40387 standard; Protein; 134 AA.
XX
AC ABP40387;
XX
XX 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5232.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX

```

PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2002-381255/41.  
DR N-PSDB; ABN92932.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
PT  
XX  
PS Disclosure; SEQ ID 5232; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
XX USPTO web site.

SQ Sequence 134 AA;

Query Match 31.2%; Score 5; DB 23; Length 134;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
| | | | |  
Db 89 IDFII 93

RESULT 37  
AAP60068  
ID AAP60068 standard; Protein; 146 AA.  
AC  
XX AAP60068;

DT 26-JUN-1991 (first entry)

DE Sequence of HTLV-III polypeptide 121 fusion protein.

XX AIDS; HIV; LAV; vaccine; diagnosis; antibody; fusion protein.  
XX HTLV-III.

XX Key Location/Qualifiers  
FH Region 1..41  
FT Protein /note= "E.coli encoded"  
FT 42..122  
FT Region /label= polypeptide 121  
FT 123..146  
FT /note= "vector encoded"

XX EPI99438-A.

XX 29-OCT-1986.

XX 03-MAR-1986; 86EP-0301488.

XX 01-MAR-1985; 85US-0707066.

XX (CHAN/) CHANG T W.  
XX (CENT-) CENTOCOR INC.

XX Chang TW, Kato I, Chanda P, Chang NT, Putney SD;

DR WPI; 1986-286179/44.  
XX  
PT Human T-cell lymphotropic virus-III polypeptide 121 - useful in  
PT assays to detect AIDS and for vaccination  
XX  
PS Claim 3; Page 14; 17pp; English.

XX The HTLV-III DNA segment of clone 121 was excised from pMR100 and  
CC sequenced (AAN60081). Based upon the DNA sequence, the putative AA  
CC sequence of the HTLV-III polypeptide could be assigned (AAP60067). In  
CC order to enhance expression the HTLV-III polypeptide in E.coli the  
CC HTLV-III segment of clone 121 was cloned into a high expression  
CC vector. E.coli transformed with the recombinant vector expressed a  
CC 15 kd fusion protein with short fusion partners at both ends (41 AAs  
CC of E.coli, the 83 AA residues encoding by HTLV-III polypeptide and  
CC 13 AA residues encoded by a multiple cloning site of the vector),  
CC (AAP60068).

SQ Sequence 146 AA;

Query Match 31.2%; Score 5; DB 7; Length 146;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
| | | | |  
Db 26 NCGID 30

RESULT 38  
AAB92462  
ID AAB92462 standard; Protein; 148 AA.  
XX  
AC AAB92462;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:10515.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.

PN EPI074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -

XX Claim 8; SEQ ID 10515; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of



the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Query Match 31.2%; Score 5; DB 22; Length 148;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QANCG 6  
Db 117 QANCG 121

RESULT 39  
AAH85902  
ID AAY85902 standard; Protein; 149 AA.

XX AAY85902;  
XX 10-APR-2000 (first entry)  
DT S. pneumoniae derived protein #111.

DE Treatment; prevention; disease; diagnosis; gene therapy; screening;  
XX bacterial; antimicrobial; antibiotic; pathogenesis; infection.

OS Streptococcus pneumoniae.

XX WO9806734-A1.

PN 19-FEB-1998.

PD 15-AUG-1997; 97WO-US14436.

PF 16-AUG-1996; 96US-0024022.

PR (SMIK ) SMITHKLINE BEECHAM CORP.

XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;

PI Stodola RK;

XX WPI; 1998-159452/14.

DR N-PSDB; AAZ96272.

XX Streptococcus pneumoniae proteins and related DNA - useful for  
PT screening compounds for antibacterial activity

XX Claim 5; Page 396; 640pp; English.

PS This invention describes novel isolated Streptococcus pneumoniae  
XX polynucleotides (see AAZ96173-296494) and their encoded proteins (see  
CC AAY85792-Y86182). The DNA, vectors and host cells described in the  
CC method of the invention are useful for the recombinant expression of the  
CC polypeptides. The polypeptides are useful for treatment or prevention of  
CC disease, or diagnosis of disease related to expression or activity of  
CC such a polypeptide. They can also be used to screen for compounds which

CC interact with and inhibit or activate such a polypeptide. The  
CC polypeptides (or DNA encoding them, via gene therapy) are also useful  
CC for inducing an immunological response in a mammal. The antagonists are  
CC useful to inhibit such bacterial polypeptides. The polypeptides are  
CC particularly useful to identify antimicrobial compounds and antibiotics.  
CC They are also useful to determine their role in pathogenesis of  
CC infection, dysfunction and disease.

XX Sequence 149 AA;

Query Match 31.2%; Score 5; DB 19; Length 149;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11  
Db 80 IDFI 84

RESULT 40  
AAW20434  
ID AAW20434 standard; protein; 150 AA.

XX AAW20434;

XX 14-JUL-1997 (first entry)

DE H. pylori cytoplasmic protein, 34189716.aa.

XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope;  
KW amino acid; metabolism.

XX Helicobacter pylori.

XX WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

XX 01-APR-1996; 96US-0630405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTR ) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaard BL;

XX WPI; 1997-052306/05.

XX N-PSDB; AAT67608.

XX Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter

XX Claim 61; Pages 610-611; 1481pp; English.

PS The present sequence is a Helicobacter pylori cytoplasmic protein  
XX involved in amino acid metabolism. The protein may be used in a vaccine  
CC to prevent or treat H. pylori infection or to identify H. pylori  
CC polypeptide binding compounds, useful as potential H. pylori life cycle  
CC activators or inhibitors. The genomic sequence of H. pylori (ATCC 55679)  
CC was determined from overlapping contigs generated by mechanically  
CC shearing the bacterial DNA. The sequences were analysed for ORF of at  
CC least 180 nucleotides, and the predicted coding regions defined by  
CC computer evaluation. To identify likely H. pylori antigens for vaccine  
CC development, the amino acid sequences predicted from various ORF were  
CC analysed for significant homology to other known or exported membrane  
CC proteins. Having identified and determined the sequences of interest,  
CC particular regions can be isolated from H. pylori by PCR amplification  
CC for recombinant polypeptide production, e.g. in E. coli hosts.

XX

SQ Sequence 150 AA;

## Query Match

Best Local Similarity 31.2%; Score 5; DB 18; Length 150;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 111 GIDFI 115

## RESULT 41

ABG30352  
ID ABG30352 standard; Protein; 151 AA.

XX

AC ABG30352;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #30343.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

XX

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

XX

DR N-PSDB; AAS94539.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX

PS Claim 20; SEQ ID No 60711; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 151 AA;

## Query Match

Best Local Similarity 31.2%; Score 5; DB 22; Length 151;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGIDF 9  
Db 144 CGIDF 148

## RESULT 42

ABG22986  
ID ABG22986 standard; Protein; 152 AA.

XX

AC ABG22986;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #22977.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

XX

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

XX

DR N-PSDB; AAS87173.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX

PS Claim 20; SEQ ID No 53345; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 152 AA;

Matches	5;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	12	FWIFW 16							
Db	115	FWIFW 119							
RESULT 43									
ABP29925	ID	ABP29925 standard; Protein; 157 AA.							
XX	AC	ABP29925;							
XX	DT	02-JUL-2002 (first entry)							
XX	DE	Streptococcus polypeptide SEQ ID NO 9026.							
XX	KW	Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;							
XX	KW	group A streptococcus; Streptococcus pyogenes; antibacterial;							
XX	KW	antiinflammatory; infection; vaccine; meningitis; gene therapy.							
XX	OS	Streptococcus agalactiae.							
XX	PN	WO200234771-A2.							
XX	PD	02-MAY-2002.							
XX	PF	29-OCT-2001; 2001WO-GB04789.							
XX	PR	27-OCT-2000; 2000GB-0026333.							
XX	PR	24-NOV-2000; 2000GB-0028727.							
XX	PR	07-MAR-2001; 2001GB-0005640.							
XX	PA	(CHIR-) CHIRON SPA.							
XX	PA	(GENO-) INST GENOMIC RES.							
XX	PI	Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;							
XX	PI	Tettelin H;							
XX	DR	WPI; 2002-352536/38.							
XX	DR	N-PSDB; ABN70556.							
XX	PT	New Streptococcus protein for the treatment or prevention of infection							
XX	PT	or disease caused by Streptococcus bacteria, such as meningitis, and							
XX	PT	for detecting a compound that binds to the protein -							
XX	PS	Claim 1; Page 4024; 4525pp; English.							
XX	CC	The invention relates to a protein (ABP25413-ABP30895) from group B							
XX	CC	streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS							
XX	CC	(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in							
XX	CC	the specification. The proteins have antibacterial and antiinflammatory							
XX	CC	activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and							
XX	CC	antibodies that bind (I) are used in the manufacture of medicaments for							
XX	CC	the treatment or prevention of infection or disease caused by							
XX	CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.							
XX	CC	Nucleic acids encoding (I) are used to detect Streptococcus in a							
XX	CC	biological sample. (I) is used to determine whether a compound binds to							
XX	CC	(I). A composition comprising (I) or a nucleic acid encoding (I), may be							
XX	CC	used as a vaccine or diagnostic composition. The disease caused by							
XX	CC	Streptococcus that is prevented or treated may be meningitis. Nucleic							
XX	CC	acid encoding (I) may be used to recombinantly produce (I) and may be							
XX	CC	used in gene therapy. Antibodies to (I) are used for affinity							
XX	CC	chromatography, immunoassays, and distinguishing/identifying							
XX	CC	Streptococcus proteins.							
XX	SQ	Sequence 157 AA;							
Query Match 31.2%; Score 5; DB 23; Length 157;									
Best Local Similarity 100.0%; Pred. No. 2.1e+02;									
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									

Db	34	IDFII 38	
RESULT 44			
ABP30763	ID	ABP30763 standard; Protein; 157 AA.	
XX	AC	ABP30763;	
XX	DT	02-JUL-2002 (first entry)	
XX	DE	Streptococcus polypeptide SEQ ID NO 10702.	
XX	KW	Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;	
XX	KW	group A streptococcus; Streptococcus pyogenes; antibacterial;	
XX	KW	antiinflammatory; infection; vaccine; meningitis; gene therapy.	
XX	OS	Streptococcus agalactiae.	
XX	PN	WO200234771-A2.	
XX	PD	02-MAY-2002.	
XX	PF	29-OCT-2001; 2001WO-GB04789.	
XX	PR	27-OCT-2000; 2000GB-0026333.	
XX	PR	24-NOV-2000; 2000GB-0028727.	
XX	PR	07-MAR-2001; 2001GB-0005640.	
XX	PA	(CHIR-) CHIRON SPA.	
XX	PA	(GENO-) INST GENOMIC RES.	
XX	PI	Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;	
XX	PI	Tettelin H;	
XX	DR	WPI; 2002-352536/38.	
XX	DR	N-PSDB; ABN71394.	
XX	PT	New Streptococcus protein for the treatment or prevention of infection	
XX	PT	or disease caused by Streptococcus bacteria, such as meningitis, and	
XX	PT	for detecting a compound that binds to the protein -	
XX	PS	Claim 1; Page 4179; 4525pp; English.	
XX	CC	The invention relates to a protein (ABP25413-ABP30895) from group B	
XX	CC	streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS	
XX	CC	(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in	
XX	CC	the specification. The proteins have antibacterial and antiinflammatory	
XX	CC	activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and	
XX	CC	antibodies that bind (I) are used in the manufacture of medicaments for	
XX	CC	the treatment or prevention of infection or disease caused by	
XX	CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.	
XX	CC	Nucleic acids encoding (I) are used to detect Streptococcus in a	
XX	CC	biological sample. (I) is used to determine whether a compound binds to	
XX	CC	(I). A composition comprising (I) or a nucleic acid encoding (I), may be	
XX	CC	used as a vaccine or diagnostic composition. The disease caused by	
XX	CC	Streptococcus that is prevented or treated may be meningitis. Nucleic	
XX	CC	acid encoding (I) may be used to recombinantly produce (I) and may be	
XX	CC	used in gene therapy. Antibodies to (I) are used for affinity	
XX	CC	chromatography, immunoassays, and distinguishing/identifying	
XX	CC	Streptococcus proteins.	
XX	SQ	Sequence 157 AA;	
Query Match 31.2%; Score 5; DB 23; Length 157;			
Best Local Similarity 100.0%; Pred. No. 2.1e+02;			
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			

Query Match 31.2%; Score 5; DB 23; Length 157;  
Best Local Similarity 100.0%; Pred. No. 2.le+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11

Query Match 31.2%; Score 5; DB 23; Length 157;  
Best Local Similarity 100.0%; Pred. No. 2.le+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
Db 34 IDFII 38

Db 34 IDFII 38

RESULT 44  
ABP30763  
ID ABP30763 standard; Protein; 157 AA.

XX AC ABP30763;

XX AC ABP30763;

XX DT 02-JUL-2002 (first entry)

XX DE Streptococcus polypeptide SEQ ID NO 10702.

XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;

XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX OS Streptococcus agalactiae.

XX PN WO200234771-A2.

XX PD 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-GB04789.

XX PR 27-OCT-2000; 2000GB-0026333.

XX PR 24-NOV-2000; 2000GB-0028727.

XX PR 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;

XX Tettelin H;

XX WPI; 2002-352536/38.

XX N-PSDB; ABN71394.

New Streptococcus protein for the treatment or prevention of infection  
or disease caused by Streptococcus bacteria, such as meningitis, and  
for detecting a compound that binds to the protein -

Claim 1; Page 4179; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B  
streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
the specification. The proteins have antibacterial and antiinflammatory  
activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
antibodies that bind (I) are used in the manufacture of medicaments for  
the treatment or prevention of infection or disease caused by  
Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
Nucleic acids encoding (I) are used to detect Streptococcus in a  
biological sample. (I) is used to determine whether a compound binds to  
(I). A composition comprising (I) or a nucleic acid encoding (I), may be  
used as a vaccine or diagnostic composition. The disease caused by  
Streptococcus that is prevented or treated may be meningitis. Nucleic  
acid encoding (I) may be used to recombinantly produce (I) and may be  
used in gene therapy. Antibodies to (I) are used for affinity  
chromatography, immunoassays, and distinguishing/identifying  
Streptococcus proteins.

Sequence 157 AA;

Job time : 11.3451 secs

RESULT 45  
ABG12807  
ID ABG12807 standard; Protein; 160 AA.  
XX  
AC ABG12807;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #12798.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS76994.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 20; SEQ ID No 43166; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 160 AA;

Query Match  
Best Local Similarity 31.2%; Score 5; DB 22; Length 160;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 10 IIFWI 14  
Db 139 IIFWI 143

Search completed: May 11, 2003, 20:10:50



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: May 11, 2003, 20:09:32 ; Search time 3.9646 Seconds  
(without alignments)  
118.743 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	37.5	399	1	US-08-414-926A-5
2	6	37.5	399	2	US-08-926-922-5
3	6	37.5	399	3	US-09-253-682-5
4	6	37.5	399	4	US-09-527-657-5
5	5	31.2	69	1	US-08-410-804-5
6	5	31.2	69	1	US-08-259-514-5
7	5	31.2	69	2	US-08-858-311-5
8	5	31.2	74	4	US-09-134-001C-5666
9	5	31.2	103	2	US-08-529-878B-41
10	5	31.2	118	4	US-09-627-376-17
11	5	31.2	134	4	US-09-134-001C-5232
12	5	31.2	220	3	US-08-228-208A-21
13	5	31.2	223	1	US-08-505-058-5
14	5	31.2	223	2	US-08-459-818-25
15	5	31.2	223	2	US-08-889-666-25
16	5	31.2	223	2	US-08-465-078-25
17	5	31.2	223	2	US-08-725-776-25
18	5	31.2	223	2	US-08-488-062-25
19	5	31.2	229	4	US-09-134-001C-2965
20	5	31.2	238	4	US-08-858-207A-420
21	5	31.2	270	3	US-09-188-579-84
22	5	31.2	270	4	US-09-315-444-84
23	5	31.2	270	4	US-09-721-362-84
24	5	31.2	295	1	US-08-118-270-79
25	5	31.2	295	5	PCT-US93-08528-79
26	5	31.2	401	3	US-08-974-022-2
27	5	31.2	401	4	US-08-795-445A-2

28	5	31.2	401	4	US-08-795-447A-2	Sequence 2, Appli
29	5	31.2	401	4	US-08-974-186-2	Sequence 2, Appli
30	5	31.2	401	4	US-08-795-446B-2	Sequence 2, Appli
31	5	31.2	401	4	US-08-706-945D-124	Sequence 124, App
32	5	31.2	518	4	US-09-134-001C-4451	Sequence 4451, Ap
33	5	31.2	580	4	US-08-913-159-10	Sequence 10, Appl
34	5	31.2	589	1	US-08-399-646-2	Sequence 2, Appli
35	5	31.2	589	1	US-08-607-321-2	Sequence 2, Appli
36	5	31.2	589	2	US-08-961-240-2	Sequence 2, Appli
37	5	31.2	589	2	US-08-605-501-2	Sequence 2, Appli
38	5	31.2	596	1	US-08-399-646-12	Sequence 12, Appl
39	5	31.2	596	1	US-08-607-321-12	Sequence 12, Appl
40	5	31.2	596	2	US-08-961-240-12	Sequence 12, Appl
41	5	31.2	596	2	US-08-605-501-12	Sequence 12, Appl
42	5	31.2	600	6	5268463-2	Patent No. 5268463
43	5	31.2	602	2	US-08-882-704A-5	Sequence 5, Appli
44	5	31.2	602	4	US-09-151-957-5	Sequence 5, Appli
45	5	31.2	602	6	5432081-2	Patent No. 5432081
46	5	31.2	603	4	US-09-149-727-6	Sequence 6, Appli
47	5	31.2	832	3	US-08-630-820-7	Sequence 7, Appli
48	5	31.2	1242	4	US-09-488-270A-2	Sequence 2, Appli
49	5	31.2	1724	2	US-08-477-451-15	Sequence 15, Appl
50	5	31.2	2763	3	US-08-496-944-2	Sequence 2, Appli
51	4	25.0	8	4	US-09-187-859-1001	Sequence 1001, Ap
52	4	25.0	8	4	US-09-187-859-1322	Sequence 1322, Ap
53	4	25.0	8	4	US-09-187-859-1903	Sequence 1903, Ap
54	4	25.0	8	4	US-09-187-859-3225	Sequence 3225, Ap
55	4	25.0	9	4	US-09-187-859-2885	Sequence 2885, Ap
56	4	25.0	9	4	US-09-187-859-3226	Sequence 3226, Ap
57	4	25.0	9	4	US-09-187-859-3897	Sequence 3897, Ap
58	4	25.0	9	4	US-09-187-859-3963	Sequence 3963, Ap
59	4	25.0	10	4	US-09-187-859-3010	Sequence 3010, Ap
60	4	25.0	10	4	US-09-187-859-3227	Sequence 3227, Ap
61	4	25.0	10	4	US-09-187-859-3902	Sequence 3902, Ap
62	4	25.0	10	4	US-09-187-859-3968	Sequence 3968, Ap
63	4	25.0	13	4	US-09-258-754-108	Sequence 108, App
64	4	25.0	13	4	US-09-042-107-108	Sequence 108, App
65	4	25.0	15	4	US-09-009-953-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1  
US-08-414-926A-5  
; Sequence 5, Application US/08414926A  
; Patent No. 5721354  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; APPLICANT: Cha, Tai-An  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum  
; STREET: 5 Palo Alto Square  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-2155  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/414,926A  
; FILING DATE: March 31, 1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: AVIR-011/OOUS  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-494-7622  
; TELEFAX: 415-857-0663  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 399 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-414-926A-5

Query Match 37.5%; Score 6; DB 1; Length 399;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFWI 14  
| | | | |  
Db 20 FIIFWI 25

RESULT 2  
US-08-926-922-5  
; Sequence 5, Application US/08926922  
; Patent No. 5925751  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; APPLICANT: Cha, Tai-An  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Luann Cserr Attorney at Law  
; STREET: 750 Arimo Avenue  
; CITY: Oakland  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94610

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/926,922  
; FILING DATE: September 10, 1997  
; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: AVIR 11A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-834-1448  
; TELEFAX: 510-839-7810  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 399 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-926-922-5

Query Match 37.5%; Score 6; DB 2; Length 399;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFWI 14  
| | | | |  
Db 20 FIIFWI 25

RESULT 3  
US-09-253-682-5  
; Sequence 5, Application US/09253682  
; Patent No. 6040170  
; GENERAL INFORMATION:

; APPLICANT: Spaete, Richard  
; APPLICANT: Cha, Tai-An  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Luann Cserr Attorney at Law  
; STREET: 750 Arimo Avenue  
; CITY: Oakland  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94610  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/253,682  
; FILING DATE:

; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/926,922  
; FILING DATE: September 10, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: AVIR 11A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-834-1448  
; TELEFAX: 510-839-7810  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 399 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-253-682-5

Query Match 37.5%; Score 6; DB 3; Length 399;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFWI 14  
| | | | |  
Db 20 FIIFWI 25

RESULT 4  
US-09-527-657-5  
; Sequence 5, Application US/09527657  
; Patent No. 6291236  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; Cha, Tai-An  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Luann Cserr Attorney at Law  
; STREET: 750 Arimo Avenue  
; CITY: Oakland  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94610

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/527,657  
; FILING DATE: 17-Mar-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:

;/ APPLICATION NUMBER: US/08/926,922  
;/ FILING DATE: September 10, 1997  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Cseri, Luann  
;/ REGISTRATION NUMBER: 31,822  
;/ REFERENCE/DOCKET NUMBER: AVIR 11A  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: 510-834-1448  
;/ TELEFAX: 510-839-7810  
;/ INFORMATION FOR SEQ ID NO: 5:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 399 amino acids  
;/ TYPE: amino acid  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: protein  
;/ SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-527-657-5

Query Match 37.5%; Score 6; DB 4; Length 399;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFWI 14  
Db 20 FIIFWI 25

RESULT 5  
US-08-410-804-5  
; Sequence 5, Application US/08410804  
; Patent No. 5632994  
; GENERAL INFORMATION:  
; APPLICANT: Reed, John C.  
; APPLICANT: Sato, Takaaki  
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cathryn Campbell  
; STREET: 4370 La Jolla Village Drive. Ste 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/410,804  
; FILING DATE: 27-MAR-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/259,514  
; FILING DATE: 14-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 1389  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 69 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-410-804-5

Query Match 31.2%; Score 5; DB 1; Length 69;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
Db 58 IDFII 62  
RESULT 6  
US-08-259-514-5  
; Sequence 5, Application US/08259514  
; Patent No. 5747245  
; GENERAL INFORMATION:  
; APPLICANT: Reed, John C.  
; APPLICANT: Sato, Takaaki  
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cathryn Campbell  
; STREET: 4370 La Jolla Village Drive. Ste 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/259,514  
; FILING DATE: 14-JUN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 9954  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 69 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-259-514-5

Query Match 31.2%; Score 5; DB 1; Length 69;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
Db 58 IDFII 62

RESULT 7  
US-08-858-311-5  
; Sequence 5, Application US/08858311  
; Patent No. 5876939  
; GENERAL INFORMATION:  
; APPLICANT: Reed, John C.  
; APPLICANT: Sato, Takaaki  
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cathryn Campbell  
; STREET: 4370 La Jolla Village Drive. Ste 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/858,311  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/410,804  
;; FILING DATE: 27-MAR-1995  
;; APPLICATION NUMBER: US 08/259,514  
;; FILING DATE: 14-JUN-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Campbell, Cathryn  
;; REGISTRATION NUMBER: 31,815  
;; REFERENCE/DOCKET NUMBER: P-LJ 1389  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (619) 535-9001  
;; TELEFAX: (619) 535-8949  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 69 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-858-311-5

Query Match 31.2%; Score 5; DB 2; Length 69;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11  
|1111  
Db 58 IDFI 62

RESULT 8  
US-09-134-001C-5666  
; Sequence 5666, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5666  
; LENGTH: 74  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5666

Query Match 31.2%; Score 5; DB 4; Length 74;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|1111  
Db 40 FIIFW 44

RESULT 9  
US-08-529-878B-41  
; Sequence 41, Application US/08529878B  
; Patent No. 5932556  
; GENERAL INFORMATION:  
; APPLICANT: Tam, Robert C.

;; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
;; TITLE OF INVENTION: REGULATION OF CD28 EXPRESSION  
;; NUMBER OF SEQUENCES: 48  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Crockett & Fish  
;; STREET: 3000 S. Augusta Court  
;; CITY: La Habra  
;; STATE: California  
;; COUNTRY: United States of America  
;; ZIP: 90631  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: WordPerfect 6.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/529,878B  
;; FILING DATE: 13-SEP-1995  
;; CLASSIFICATION: 424  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fish, Robert D.  
;; REGISTRATION NUMBER: 33,880  
;; REFERENCE/DOCKET NUMBER: 213/003  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 714-525-3433  
;; TELEFAX: 714-525-3303  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 41:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 103 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-529-878B-41

Query Match 31.2%; Score 5; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|1111  
Db 57 FIIFW 61

RESULT 10  
US-09-627-376-17  
; Sequence 17, Application US/09627376  
; Patent No. 6342385  
; GENERAL INFORMATION:  
; APPLICANT: Q1, Fengxia Caulfield, Page Chen, Ping  
; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS  
; FILE REFERENCE: UAB-17402/22  
; CURRENT APPLICATION NUMBER: US/09/627,376  
; CURRENT FILING DATE: 2001-05-30  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 17  
; LENGTH: 118  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-627-376-17

Query Match 31.2%; Score 5; DB 4; Length 118;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 IFWIF 15  
|1111  
Db 14 IFWIF 18

RESULT 11  
US-09-134-001C-5232



; Sequence 5232, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5232  
; LENGTH: 134  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
; US-09-134-001C-5232

Query Match 31.2%; Score 5; DB 4; Length 134;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11  
Db 89 IDFI 93

RESULT 12  
US-08-228-208A-21  
; Sequence 21, Application US/08228208A  
; Patent No. 6090914  
; GENERAL INFORMATION:  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Damle, Nitin K.  
; APPLICANT: Brady, William  
; APPLICANT: Wallace, Philip M.  
; TITLE OF INVENTION: CTIA4/CD28Ig HYBRID FUSION  
; TITLE OF INVENTION: PROTEINS AND USES THEREOF  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 11150 Santa Monica Boulevard, Suite 400  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/228,208A  
; FILING DATE: 15-APR-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/008,898  
; FILING DATE: 22-JAN-1993  
; APPLICATION NUMBER: 07/723,617  
; FILING DATE: 27-JUN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Adriano, Sarah B  
; REGISTRATION NUMBER: 34,470  
; REFERENCE/DOCKET NUMBER: 30436-30US01  
; TELEPHONE: 310 445-1140  
; TELEFAX: 310 445-9031  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 220 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-228-208A-21

Query Match 31.2%; Score 5; DB 3; Length 220;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 174 FIIFW 178

RESULT 13  
US-08-505-058-5  
; Sequence 5, Application US/08505058  
; Patent No. 5773253  
; GENERAL INFORMATION:  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Peach, Robert  
; TITLE OF INVENTION: CTIA4 Mutant Molecules and Uses Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 11150 Santa Monica Blvd., Suite 400  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/505,058  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/228,208  
; FILING DATE: 15-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Adriano, Sarah B.  
; REGISTRATION NUMBER: 34,470  
; REFERENCE/DOCKET NUMBER: 30436.30US11  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 310-445-1140  
; TELEFAX: 310-445-9031  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-505-058-5

Query Match 31.2%; Score 5; DB 1; Length 223;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 176 FIIFW 180

RESULT 14  
US-08-459-818-25  
; Sequence 25, Application US/08459818

```
; Patent No. 5851795
; GENERAL INFORMATION:
; APPLICANT: Linsley, Peter S.
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Damle, Nitin K.
; APPLICANT: Brady, William
; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 11150 Santa Monica Blvd., Suite 400
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FastSeq 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,818
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Adriano, Sarah B.
; REGISTRATION NUMBER: 34,470
; REFERENCE/DOCKET NUMBER: 30436.35US02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310-445-1140
; TELEFAX: 310-445-9031
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-459-818-25

Query Match 31.2%; Score 5; DB 2; Length 223;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13
Db 176 FIIFW 180

RESULT 15
US-08-889-666-25
; Sequence 25, Application US/08889666
; Patent No. 5885579
; GENERAL INFORMATION:
; APPLICANT: Linsley, Peter S.
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Damle, Nitin K.
; APPLICANT: Brady, William
; APPLICANT: Kiener, Peter A.
; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 11150 Santa Monica Blvd., Suite 400
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/889,666
; FILING DATE: 08-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/375390
; FILING DATE: 18-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Adriano, Sarah B.
; REGISTRATION NUMBER: 34,470
; REFERENCE/DOCKET NUMBER: 30436-35US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310-445-1140
; TELEFAX: 310-445-9031
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-889-666-25

Query Match 31.2%; Score 5; DB 2; Length 223;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13
Db 176 FIIFW 180

RESULT 16
US-08-465-078-25
; Sequence 25, Application US/08465078
; Patent No. 5885796
; GENERAL INFORMATION:
; APPLICANT: Linsley, Peter S.
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Damle, Nitin K.
; APPLICANT: Brady, William
; APPLICANT: Kiener, Peter A.
; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 11150 Santa Monica Blvd., Suite 400
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,078
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/375390
; FILING DATE: 18-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adriano, Sarah B.
; REGISTRATION NUMBER: 34,470
; REFERENCE/DOCKET NUMBER: 30436-35US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310-445-1140
; TELEFAX: 310-445-9031
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
```

; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-465-078-25

Query Match 31.2%; Score 5; DB 2; Length 223;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|  
|  
|  
|  
Db 176 FIIFW 180

## RESULT 17

US-08-725-776-25  
; Sequence 25, Application US/08725776  
; Patent No. 5968510

## ; GENERAL INFORMATION:

; APPLICANT: Linsley, Peter S.  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Damle, Nitin K.  
; APPLICANT: Brady, William  
; APPLICANT: Kiener, Peter A.

; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof  
; NUMBER OF SEQUENCES: 26

## ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merchant & Gould  
; STREET: 11150 Santa Monica Blvd., Suite 400  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90025

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/725,776  
; FILING DATE:

## ; CLASSIFICATION:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/375390  
; FILING DATE: 18-JAN-1995

## ; ATTORNEY/AGENT INFORMATION:

; NAME: Adriano, Sarah B.  
; REGISTRATION NUMBER: 34,470  
; REFERENCE/DOCKET NUMBER: 30436-35US01  
; TELEPHONE: 310-445-1140  
; TELEFAX: 310-445-9031

## ; INFORMATION FOR SEQ ID NO: 25:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-725-776-25

Query Match 31.2%; Score 5; DB 2; Length 223;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|  
|  
|  
|  
Db 176 FIIFW 180

## RESULT 18

US-08-488-062-25  
; Sequence 25, Application US/08488062  
; Patent No. 5977318

## ; GENERAL INFORMATION:

; APPLICANT: Linsley, Peter S.  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Damle, Nitin K.  
; APPLICANT: Brady, William  
; APPLICANT: Kiener, Peter A.

; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof  
; NUMBER OF SEQUENCES: 26

## ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merchant & Gould  
; STREET: 11150 Santa Monica Blvd., Suite 400  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90025

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,062  
; FILING DATE: 07-JUN-1995

## ; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/375390

; FILING DATE: 18-JAN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Adriano, Sarah B.

; REGISTRATION NUMBER: 34,470

; REFERENCE/DOCKET NUMBER: 30436-35US01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 310-445-1140

; TELEFAX: 310-445-9031

; INFORMATION FOR SEQ ID NO: 25:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 223 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-488-062-25

Query Match 31.2%; Score 5; DB 2; Length 223;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|  
|  
|  
|  
Db 176 FIIFW 180

## RESULT 19

US-09-134-001C-2965  
; Sequence 2965, Application US/09134001C  
; Patent No. 6380370

## ; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS AUREUS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 2965

; LENGTH: 229

```
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2965

Query Match          31.2%; Score 5; DB 4; Length 229;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWI 14
Db 63 IIFWI 67

RESULT 20
US-08-858-207A-420
; Sequence 420, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmil, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 420:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
US-08-858-207A-420

Query Match          31.2%; Score 5; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10
Db 32 GIDFI 36

RESULT 21
US-09-188-579-84
; Sequence 84, Application US/09188579B
```

```
; Patent No. 6107040
; GENERAL INFORMATION:
; APPLICANT: Shuman, Stewart
; TITLE OF INVENTION: Pharmacological Targeting of mRNA Cap Formation
; FILE REFERENCE: D6185
; CURRENT APPLICATION NUMBER: US/09/188,579B
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 114
; SEQ ID NO 84
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Chlorella virus PBCV-1
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of RNA guanylyltransferase.
US-09-188-579-84

Query Match          31.2%; Score 5; DB 3; Length 270;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11
Db 188 IDFII 192

RESULT 22
US-09-315-444-84
; Sequence 84, Application US/09315444A
; Patent No. 6232070
; GENERAL INFORMATION:
; APPLICANT: Shuman, Stewart
; TITLE OF INVENTION: Pharmacological Targeting of mRNA Cap Formation
; FILE REFERENCE: D6185CIP
; CURRENT APPLICATION NUMBER: US/09/315,444A
; CURRENT FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: US 09/188,579
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 84
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Chlorella virus PBCV-1
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of RNA guanylyltransferase.
US-09-315-444-84

Query Match          31.2%; Score 5; DB 4; Length 270;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11
Db 188 IDFII 192

RESULT 23
US-09-721-362-84
; Sequence 84, Application US/09721362
; Patent No. 6420163
; GENERAL INFORMATION:
; APPLICANT: Shuman, Stewart
; TITLE OF INVENTION: Pharmacological Targeting of mRNA Cap Formation
; FILE REFERENCE: D6185CIP/D
; CURRENT APPLICATION NUMBER: US/09/721,362
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: US 09/315,444
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 84
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Chlorella virus PBCV-1
; FEATURE:
```



; OTHER INFORMATION: Amino acid sequence of RNA guanylyltransferase.  
US-09-721-362-84

Query Match 31.2%; Score 5; DB 4; Length 270;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11  
Db 188 IDFI 192

RESULT 24

US-08-118-270-79  
; Sequence 79, Application US/08118270  
; Patent No. 5508384  
; GENERAL INFORMATION:  
; APPLICANT: Murphy, Randall B.  
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
; NUMBER OF SEQUENCES: 348  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/118,270  
; FILING DATE: 09-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/943,236  
; FILING DATE: 10-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Townsend, Kevin G.  
; REGISTRATION NUMBER: 34,033  
; REFERENCE/DOCKET NUMBER: MURPHY=2A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633

INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 295 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US93-08528-79

Query Match 31.2%; Score 5; DB 1; Length 295;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 207 FIIFW 211

RESULT 25

PCT-US93-08528-79  
; Sequence 79, Application PC/TUS9308528  
; GENERAL INFORMATION:  
; APPLICANT: New York University  
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF

US-08-118-270-79

Query Match 31.2%; Score 5; DB 1; Length 295;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 207 FIIFW 211

Query Match 31.2%; Score 5; DB 5; Length 295;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 207 FIIFW 211

RESULT 26

US-08-974-022-2  
; Sequence 2, Application US/08974022  
; Patent No. 6015938  
; GENERAL INFORMATION:  
; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.  
; APPLICANT: Calzone, Frank J.  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: OSTEOPROTEGERIN  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 Dehavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/974,022  
; FILING DATE: 12-DEC-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:

PCT-US93-08528-79

Query Match 31.2%; Score 5; DB 5; Length 295;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 207 FIIFW 211

; APPLICATION NUMBER: 08/577,788  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-378  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 401 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-974-022-2

Query Match 31.2%; Score 5; DB 3; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
|||||  
Db 194 NCGID 198

## RESULT 27

US-08-795-445A-2  
; Sequence 2, Application US/08795445A  
; Patent No. 6284485

; GENERAL INFORMATION:  
; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.  
; APPLICANT: Calzone, Frank J.  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: OSTEOPROTEGERIN  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 Dehavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/795,445A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/577,788  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-378  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 401 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-795-445A-2

Query Match 31.2%; Score 5; DB 4; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
|||||  
Db 194 NCGID 198

## RESULT 28

US-08-795-447A-2  
; Sequence 2, Application US/08795447A  
; Patent No. 6284728  
; GENERAL INFORMATION:  
; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.  
; APPLICANT: Calzone, Frank J.  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: Osteoprotegerin  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: One Amgen Center Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91362-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/795,447A  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-378D2  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 401 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-795-447A-2

Query Match 31.2%; Score 5; DB 4; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
|||||  
Db 194 NCGID 198

## RESULT 29

US-08-974-186-2  
; Sequence 2, Application US/08974186  
; Patent No. 6284740

; GENERAL INFORMATION:  
; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.  
; APPLICANT: Calzone, Frank J.  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: OSTEOPROTEGERIN  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 Dehavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/974,186  
; FILING DATE:  
; CLASSIFICATION:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/577,788  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-378  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 401 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-974-186-2

Query Match 31.2%; Score 5; DB 4; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
Db 194 NCGID 198

RESULT 30  
US-08-795-446B-2  
; Sequence 2, Application US/08795446B  
; Patent No. 6288032  
; GENERAL INFORMATION:  
; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.  
; APPLICANT: Calzone, Frank J.  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: OSTEOPROTEGERIN  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 Dehavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/795,446B  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/577,788  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-378  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 401 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-795-446B-2

Query Match 31.2%; Score 5; DB 4; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
Db 194 NCGID 198

RESULT 31  
US-08-706-945D-124  
; Sequence 124, Application US/08706945D  
; Patent No. 6369027  
; GENERAL INFORMATION:  
; APPLICANT: Boyle, William  
; APPLICANT: Lacey, David  
; APPLICANT: Calzone, Frank  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: Osteoprotegerin  
; FILE REFERENCE: A-378CIP  
; CURRENT APPLICATION NUMBER: US/08/706,945D  
; CURRENT FILING DATE: 1996-09-03  
; PRIOR APPLICATION NUMBER: 08/577,788  
; PRIOR FILING DATE: 1995-12-22  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 124  
; LENGTH: 401  
; TYPE: PRT  
; ORGANISM: Rattus rattus  
US-08-706-945D-124

Query Match 31.2%; Score 5; DB 4; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
Db 194 NCGID 198

RESULT 32  
US-09-134-001C-4451  
; Sequence 4451, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4451  
; LENGTH: 518  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4451

Query Match 31.2%; Score 5; DB 4; Length 518;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 141 GIDFI 145

RESULT 33  
US-08-913-159-10  
; Sequence 10, Application US/08913159  
; Patent No. 6300109  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Plasmid-derived type II  
; TITLE OF INVENTION: restriction-modification systems from Lactococcus lactis  
; NUMBER OF SEQUENCES: 14  
; COMPUTER READABLE FORM:

```
;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,159
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 0179/95
; FILING DATE: 17-FEB-1995
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-913-159-10
```

```
Query Match 31.2%; Score 5; DB 4; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 7 IDFI 11
      |||||
Db 148 IDFI 152
```

```
RESULT 34
US-08-399-646-2
; Sequence 2, Application US/08399646
; Patent No. 5556781
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: HATTORI, Kazuko
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,646
; FILING DATE: 07-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 59834
; FILING DATE: 07-MAR-1994
; APPLICATION DATA:
; FILING DATE: 07-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: KUBOTA=5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: amino acid
```

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-399-646-2
Query Match 31.2%; Score 5; DB 1; Length 589;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 6 GIDFI 10
      |||||
Db 138 GIDFI 142
```

```
RESULT 35
US-08-607-321-2
; Sequence 2, Application US/08607321
; Patent No. 5716813
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: HATTORI, Kazuko
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/607,321
; FILING DATE: 26-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,646
; FILING DATE: 07-MAR-1995
; APPLICATION NUMBER: JP 59834
; FILING DATE: 07-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 59840
; FILING DATE: 07-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: KUBOTA=5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-607-321-2
```

```
Query Match 31.2%; Score 5; DB 1; Length 589;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 6 GIDFI 10
      |||||
Db 138 GIDFI 142
```



## RESULT 36

US-08-961-240-2  
; Sequence 2, Application US/08961240  
; Patent No. 5830715  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: HATTORI, Kazuko  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,240  
; FILING DATE: 30-OCT-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,646  
; FILING DATE: 07-MAR-1995  
; APPLICATION NUMBER: JP 59834  
; FILING DATE: 07-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 59840  
; FILING DATE: 07-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: KUBOTA=5  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 589 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-961-240-2

Query Match 31.2%; Score 5; DB 2; Length 589;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 GIDFI 10  
|||||  
Db 138 GIDFI 142

## RESULT 37

US-08-605-501-2  
; Sequence 2, Application US/08605501  
; Patent No. 5834287  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: HATTORI, Kazuko  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE

; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/605,501  
; FILING DATE: 26-FEB-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,646  
; FILING DATE: 07-MAR-1995  
; APPLICATION NUMBER: JP 59834  
; FILING DATE: 07-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 59840  
; FILING DATE: 07-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: KUBOTA=5  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 589 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-605-501-2

Query Match 31.2%; Score 5; DB 2; Length 589;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 GIDFI 10  
|||||  
Db 138 GIDFI 142

## RESULT 38

US-08-399-646-12  
; Sequence 12, Application US/083999646  
; Patent No. 5556781  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: HATTORI, Kazuko  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/399,646  
;; FILING DATE: 07-MAR-1995  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 59834  
;; FILING DATE: 07-MAR-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 59840  
;; FILING DATE: 07-MAR-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BROWDY, Roger L.  
;; REGISTRATION NUMBER: 25,618  
;; REFERENCE/DOCKET NUMBER: KUBOTA-5  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-628-5197  
;; TELEFAX: 202-737-3528  
;; TELEX: 248633  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 596 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-399-646-12

Query Match 31.2%; Score 5; DB 1; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|||||  
Db 145 GIDFI 149

## RESULT 39

;; Sequence 12, Application US/08607321  
;; Patent No. 5716813  
;; GENERAL INFORMATION:  
;; APPLICANT: KUBOTA, Michio  
;; APPLICANT: TSUSAKI, Keiji  
;; APPLICANT: HATTORI, Kazuko  
;; APPLICANT: SUGIMOTO, Toshiyuki  
;; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
;; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
;; NUMBER OF SEQUENCES: 18  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: BROWDY AND NEIMARK  
;; STREET: 419 Seventh Street, N.W., Suite 300  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20004  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/607,321  
;; FILING DATE: 26-FEB-1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/399,646  
;; FILING DATE: 07-MAR-1995  
;; APPLICATION NUMBER: JP 59834  
;; FILING DATE: 07-MAR-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 59840  
;; FILING DATE: 07-MAR-1994  
;; ATTORNEY/AGENT INFORMATION:

;; NAME: BROWDY, Roger L.  
;; REGISTRATION NUMBER: 25,618  
;; REFERENCE/DOCKET NUMBER: KUBOTA-5  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-628-5197  
;; TELEFAX: 202-737-3528  
;; TELEX: 248633  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 596 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-607-321-12

Query Match 31.2%; Score 5; DB 1; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|||||  
Db 145 GIDFI 149

## RESULT 40

;; US-08-961-240-12  
;; Sequence 12, Application US/08961240  
;; Patent No. 5830715  
;; GENERAL INFORMATION:  
;; APPLICANT: KUBOTA, Michio  
;; APPLICANT: TSUSAKI, Keiji  
;; APPLICANT: HATTORI, Kazuko  
;; APPLICANT: SUGIMOTO, Toshiyuki  
;; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
;; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
;; NUMBER OF SEQUENCES: 18  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: BROWDY AND NEIMARK  
;; STREET: 419 Seventh Street, N.W., Suite 300  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20004  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/961,240  
;; FILING DATE: 30-OCT-1997  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/399,646  
;; FILING DATE: 07-MAR-1995  
;; APPLICATION NUMBER: JP 59834  
;; FILING DATE: 07-MAR-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 59840  
;; FILING DATE: 07-MAR-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BROWDY, Roger L.  
;; REGISTRATION NUMBER: 25,618  
;; REFERENCE/DOCKET NUMBER: KUBOTA-5  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-628-5197  
;; TELEFAX: 202-737-3528  
;; TELEX: 248633  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 596 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear

; MOLECULE TYPE: protein  
US-08-961-240-12

Query Match 31.2%; Score 5; DB 2; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|  
|  
|  
|  
Db 145 GIDFI 149

## RESULT 41

US-08-605-501-12  
; Sequence 12, Application US/08605501

; Patent No. 5834287

; GENERAL INFORMATION:

; APPLICANT: KUBOTA, Michio

; APPLICANT: TSUSAKI, Keiji

; APPLICANT: HATTORI, Kazuko

; APPLICANT: SUGIMOTO, Toshiyuki

; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND

; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 Seventh Street, N.W., Suite 300

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/605,501

; FILING DATE: 26-FEB-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/399,646

; FILING DATE: 07-MAR-1995

; APPLICATION NUMBER: JP 59834

; FILING DATE: 07-MAR-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 59840

; FILING DATE: 07-MAR-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: BROWDY, Roger L.

; REGISTRATION NUMBER: 25,618

; REFERENCE/DOCKET NUMBER: KUBOTA=5

; TELEPHONE: 202-628-5197

; TELEFAX: 202-737-3528

; TELEX: 248633

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 596 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-605-501-12

## Query Match

31.2%; Score 5; DB 2; Length 596;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10

|  
|  
|  
|

Db 145 GIDFI 149

## RESULT 42

5268463-2

; Patent No. 5268463

; APPLICANT: JEFFERSON, RICHARD A.

; TITLE OF INVENTION: PLAANT PROMOTER a-GLUCURONIDASE GENE

; CONSTRUCT

; NUMBER OF SEQUENCES: 9

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/447,976

; FILING DATE: 08-DEC-1989

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 119,102

; FILING DATE: 10-NOV-1987

; APPLICATION NUMBER: 264,586

; FILING DATE: 31-OCT-1988

; SEQ ID NO:2:

; LENGTH: 600

5268463-2

## Query Match

31.2%; Score 5; DB 6; Length 600;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8

|  
|  
|  
|

Db 27 NCGID 31

## RESULT 43

US-08-882-704A-5

; Sequence 5, Application US/08882704A

; Patent No. 5879906

; GENERAL INFORMATION:

; APPLICANT: Jefferson, Richard A.

; APPLICANT: Wilson, Katherine J.

; APPLICANT: Leader, Michael

; TITLE OF INVENTION: GLUCURONIDE REPRESSORS AND USES THEREOF

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/882,704A

; FILING DATE: 25-JUN-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: No. 5879906tenburg Ph.D., Carol

; REGISTRATION NUMBER: 39,317

; REFERENCE/DOCKET NUMBER: 190106.404

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 602 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-882-704A-5

## Query Match

31.2%; Score 5; DB 2; Length 602;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
| | | | |  
Db 27 NCGID 31

## RESULT 44

US-09-151-957-5

; Sequence 5, Application US/09151957

; Patent No. 6429292

; GENERAL INFORMATION:

; APPLICANT: Jefferson, Richard A.

; Wilson, Katherine J.

; Leader, Michael

; TITLE OF INVENTION: GLUCURONIDE REPRESSORS AND USES THEREOF

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/151,957

; FILING DATE: 11-Sep-1998

; CLASSIFICATION: &lt;Unknown&gt;

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/882,704

; FILING DATE: &lt;Unknown&gt;

; ATTORNEY/AGENT INFORMATION:

; NAME: No. 6429292tenburg Ph.D., Carol

; REGISTRATION NUMBER: 39,317

; REFERENCE/DOCKET NUMBER: 190106.404

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 602 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-151-957-5

## Query Match

Best Local Similarity 31.2%; Score 5; DB 4; Length 602;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
| | | | |  
Db 27 NCGID 31

## RESULT 45

5432081-2

; Patent No. 5432081

; APPLICANT: JEFFERSON, RICHARD A.

; TITLE OF INVENTION: HOST CELLS TRANSFORMED WITH THE E.COLI

; GLUCORONIDE PERMEASE GENE

; NUMBER OF SEQUENCES: 10

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/138,546

; FILING DATE: 15-OCT-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 447,976

; FILING DATE: 08-DEC-1989

; APPLICATION NUMBER: 264,586



Tue May 13 12:12:55 2003

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:12:57 ; Search time 5.23894 Seconds  
(without alignments)  
281.051 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size : 0

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	100.0	16	9	US-09-854-133-587
2	16	100.0	16	10	US-09-738-973-587
3	16	100.0	97	9	US-09-854-133-586
4	16	100.0	97	10	US-09-738-973-586
5	5	31.2	41	9	US-09-820-843A-43
6	5	31.2	44	10	US-09-864-761-39223
7	5	31.2	67	9	US-10-073-961-221
8	5	31.2	67	10	US-09-764-887-221
9	5	31.2	68	9	US-10-102-806-693
10	5	31.2	82	9	US-09-764-891-3247
11	5	31.2	104	9	US-09-764-891-3077
12	5	31.2	118	12	US-10-047-676A-17
13	5	31.2	166	9	US-09-738-626-4531
14	5	31.2	166	9	US-09-746-560A-94
15	5	31.2	166	9	US-09-746-560A-96
16	5	31.2	198	9	US-09-791-279-200
17	5	31.2	208	10	US-09-728-721-65
18	5	31.2	216	10	US-09-747-155-47
19	5	31.2	216	10	US-09-747-155-51
					Sequence 587, App
					Sequence 587, App
					Sequence 586, App
					Sequence 586, App
					Sequence 43, Appl
					Sequence 39223, A
					Sequence 221, App
					Sequence 221, App
					Sequence 693, App
					Sequence 3247, Ap
					Sequence 3077, Ap
					Sequence 17, Appl
					Sequence 4531, Ap
					Sequence 94, Appl
					Sequence 96, Appl
					Sequence 200, App
					Sequence 65, Appl
					Sequence 47, Appl
					Sequence 51, Appl

20	5	31.2	216	10	US-09-747-155-116	Sequence 116, App
21	5	31.2	220	9	US-10-107-868-25	Sequence 25, Appl
22	5	31.2	220	9	US-09-989-545-19	Sequence 19, Appl
23	5	31.2	220	9	US-10-301-056-25	Sequence 25, Appl
24	5	31.2	220	12	US-10-107-828-25	Sequence 25, Appl
25	5	31.2	220	12	US-10-107-907-25	Sequence 25, Appl
26	5	31.2	228	9	US-09-738-626-4341	Sequence 4341, Ap
27	5	31.2	305	9	US-09-738-626-3808	Sequence 3808, Ap
28	5	31.2	312	9	US-09-864-029-2	Sequence 2, Appl
29	5	31.2	312	9	US-09-864-029-4	Sequence 4, Appl
30	5	31.2	321	10	US-09-815-242-4990	Sequence 4990, Ap
31	5	31.2	321	10	US-09-815-242-10635	Sequence 10635, A
32	5	31.2	322	10	US-09-816-028A-48	Sequence 48, Appl
33	5	31.2	345	9	US-09-886-055-293	Sequence 8, Appl
34	5	31.2	345	10	US-09-886-055-293	Sequence 293, App
35	5	31.2	388	10	US-09-880-137-5	Sequence 5, Appl
36	5	31.2	388	10	US-09-880-137-6	Sequence 6, Appl
37	5	31.2	403	9	US-09-738-626-5328	Sequence 5328, Ap
38	5	31.2	403	10	US-09-852-053-4	Sequence 4, Appl
39	5	31.2	411	10	US-09-815-242-5128	Sequence 5128, Ap
40	5	31.2	430	9	US-10-058-636-2	Sequence 2, Appl
41	5	31.2	445	9	US-09-992-598-177	Sequence 177, App
42	5	31.2	445	9	US-09-989-293A-177	Sequence 177, App
43	5	31.2	445	9	US-10-063-547-32	Sequence 32, Appl
44	5	31.2	445	9	US-09-989-735-177	Sequence 177, App
45	5	31.2	445	9	US-09-990-444-177	Sequence 177, App
46	5	31.2	445	9	US-09-989-730-177	Sequence 177, App
47	5	31.2	445	9	US-09-990-436-177	Sequence 177, App
48	5	31.2	445	9	US-09-991-181-177	Sequence 177, App
49	5	31.2	445	9	US-09-993-687-177	Sequence 177, App
50	5	31.2	445	9	US-09-989-734-177	Sequence 177, App
51	5	31.2	445	9	US-09-997-653-177	Sequence 177, App
52	5	31.2	445	9	US-10-174-590-148	Sequence 148, App
53	5	31.2	445	9	US-10-176-758-148	Sequence 148, App
54	5	31.2	445	9	US-10-063-616-32	Sequence 32, Appl
55	5	31.2	445	9	US-10-175-737-148	Sequence 148, App
56	5	31.2	445	9	US-09-993-667-177	Sequence 177, App
57	5	31.2	445	9	US-10-063-502-32	Sequence 32, Appl
58	5	31.2	445	9	US-10-173-706-148	Sequence 148, App
59	5	31.2	445	9	US-10-175-738-148	Sequence 148, App
60	5	31.2	445	9	US-10-175-752-148	Sequence 148, App
61	5	31.2	445	9	US-10-176-482-148	Sequence 148, App
62	5	31.2	445	9	US-10-176-757-148	Sequence 148, App
63	5	31.2	445	9	US-10-176-913-148	Sequence 148, App
64	5	31.2	445	9	US-10-180-552-148	Sequence 148, App
65	5	31.2	445	9	US-10-180-557-148	Sequence 148, App

ALIGNMENTS

RESULT 1  
US-09-854-133-587  
; Sequence 587, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 587  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-854-133-587

Query Match 100.0%; Score 16; DB 9; Length 16;  
Best Local Similarity 100.0%; Pred. No. 6.9e-12;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANCGIDFIIFWIFW 16  
| | | | | | | | | | | | | | | |  
Db 1 FOANCGIDFIIFWIFW 16

RESULT 2  
US-09-738-973-587  
; Sequence 587, Application US/09738973  
; Patent No. US20020110563A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Indirias, Carol Yoseph  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Elliot, Mark  
; APPLICANT: Mannion, Jane  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C9  
; CURRENT APPLICATION NUMBER: US/09/738,973  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 587  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 587  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-738-973-587

Query Match 100.0%; Score 16; DB 10; Length 16;  
Best Local Similarity 100.0%; Pred. No. 6.9e-12;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANCGIDFIIFWIFW 16  
| | | | | | | | | | | | | | | |  
Db 1 FOANCGIDFIIFWIFW 16

RESULT 3  
US-09-854-133-586  
; Sequence 586, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 586  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-854-133-586

Query Match 100.0%; Score 16; DB 9; Length 97;

Best Local Similarity 100.0%; Pred. No. 3.3e-11;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANCGIDFIIFWIFW 16  
| | | | | | | | | | | | | | | |  
Db 35 FOANCGIDFIIFWIFW 50

RESULT 4  
US-09-738-973-586  
; Sequence 586, Application US/09738973  
; Patent No. US20020110563A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Indirias, Carol Yoseph  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Elliot, Mark  
; APPLICANT: Mannion, Jane  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C9  
; CURRENT APPLICATION NUMBER: US/09/738,973  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 587  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 586  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-738-973-586

Query Match 100.0%; Score 16; DB 10; Length 97;  
Best Local Similarity 100.0%; Pred. No. 3.3e-11;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANCGIDFIIFWIFW 16  
| | | | | | | | | | | | | | | |  
Db 35 FOANCGIDFIIFWIFW 50

RESULT 5  
US-09-820-843A-43  
; Sequence 43, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE  
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES  
; FILE REFERENCE: Q63915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 43  
; LENGTH: 41  
; TYPE: PRT  
; ORGANISM: C. jejuni  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: hypothetical protein Cj0344  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: gi|6967819  
US-09-820-843A-43

Query Match 31.2%; Score 5; DB 9; Length 41;  
Best Local Similarity 100.0%; Pred. No. 29;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
| | | | |  
Db 10 DFIIF 14

RESULT 6

US-09-864-761-39223  
; Sequence 39223, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 39223  
; LENGTH: 44  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC006504.1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.6  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3

; OTHER INFORMATION: SWISSPROT HIT: 000507, EVALUE 4.00e-03  
US-09-864-761-39223

Query Match .31.2%; Score 5; DB 10; Length 44;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
| | | | |  
Db 9 GIDFI 13

RESULT 7

US-10-073-961-221  
; Sequence 221, Application US/10073961  
; Publication No. US20030077602A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PALL3C1  
; CURRENT APPLICATION NUMBER: US/10/073,961  
; CURRENT FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: 09/764,887  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/225,757  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/226,868  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/216,647  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,267  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/216,880  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,270  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/251,869  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/235,834  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/234,274  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/234,223  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/228,924  
; PRIOR FILING DATE: 2000-08-30  
; PRIOR APPLICATION NUMBER: 60/224,518  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,369  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/224,519  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,964  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/241,809

; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/249,299  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/236,327  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/241,785  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/244,617  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/225,268  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,368  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/251,856  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/251,868  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/229,344  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/234,997  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: 60/229,343  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,345  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,287  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,513  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/231,413  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/229,509  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/236,367  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/237,039  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,038  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/236,370  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/236,802  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,037  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,040  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/240,960  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/239,935  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/239,937  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/241,787  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/246,474  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,532  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,216  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,210  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/226,681  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,759  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/225,213  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/227,182  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,214  
; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/235,836  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/230,438  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/215,135  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 60/225,266  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/249,218  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,208  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,213  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,212  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,207  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,245  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,244  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,217  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,211  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,215  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,264  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,214  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,297  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/232,400  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/231,242  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/232,081  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/232,080  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/231,414  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/231,244  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/233,064  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/233,063  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,397  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,399  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,401  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/241,808  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,826  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,786  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,221  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/246,475  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/231,243  
; PRIOR FILING DATE: 2000-09-08

Query Match 31.2%; Score 5; DB 9; Length 67;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 8 DFIIIF 12  
|||||  
Db 32 DFIIIF 36

RESULT 8  
US-09-764-887-221  
; Sequence 221, Application US/09764887  
; Patent No. US20020042096A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA113  
; CURRENT APPLICATION NUMBER: US/09/764,887  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 658  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 221  
; LENGTH: 67  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (8)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (18)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-887-221

Query Match 31.2%; Score 5; DB 10; Length 67;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12  
|||||  
Db 32 DFIIIF 36

RESULT 9  
US-10-102-806-693  
; Sequence 693, Application US/10102806  
; Publication No. US20030054421A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA103P1C1  
; CURRENT APPLICATION NUMBER: US/10/102,806  
; CURRENT FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: 09/925,298  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05881  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 846  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 693  
; LENGTH: 68  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-102-806-693

Query Match 31.2%; Score 5; DB 9; Length 68;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12  
|||||  
Db 47 DFIIIF 51

RESULT 10  
US-09-764-891-3247  
; Sequence 3247, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3247  
; LENGTH: 82  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (70)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-891-3247

Query Match 31.2%; Score 5; DB 9; Length 82;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QANCG 6  
|||||  
Db 6 QANCG 10

RESULT 11  
US-09-764-891-3077  
; Sequence 3077, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3077  
; LENGTH: 104  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-891-3077

Query Match 31.2%; Score 5; DB 9; Length 104;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
|||||  
Db 31 IDFII 35

RESULT 12  
US-10-047-676A-17  
; Sequence 17, Application US/10047676A  
; Patent No. US20020123105A1  
; GENERAL INFORMATION:  
; APPLICANT: Qi, Fengxia  
; APPLICANT: Caulfield, Page W.  
; APPLICANT: Chen, Ping W.  
; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS  
; FILE REFERENCE: UAB-17403/22  
; CURRENT APPLICATION NUMBER: US/10/047,676A  
; CURRENT FILING DATE: 2002-03-21  
; PRIOR APPLICATION NUMBER: US 09/627,376

; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 17  
; LENGTH: 118  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-047-676A-17

Query Match 31.2%; Score 5; DB 12; Length 118;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 IFWIF 15  
Db 14 IFWIF 18

RESULT 13  
US-09-738-626-4531  
; Sequence 4531, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4531  
; LENGTH: 166  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4531

Query Match 31.2%; Score 5; DB 9; Length 166;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 141 GIDFI 145

RESULT 14  
US-09-746-660A-94  
; Sequence 94, Application US/09746660A  
; Publication No. US20030049804A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Haberhauer, Gregor  
; APPLICANT: Kim, Jun-Won  
; APPLICANT: Lee, Heung-Schick  
; APPLICANT: Hwang, Byung-Joon

; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS  
; FILE REFERENCE: BGI-121CP2  
; CURRENT APPLICATION NUMBER: US/09/746,660A  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 09/606740  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 09/603124  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: 60/142101  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: 60/148613  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/187970  
; PRIOR FILING DATE: 2000-03-09  
; PRIOR APPLICATION NUMBER: DE 19931420.9  
; PRIOR FILING DATE: 1999-07-08  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: PatentIn Vers. 2.0  
; SEQ ID NO 94  
; LENGTH: 166  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-746-660A-94

Query Match 31.2%; Score 5; DB 9; Length 166;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 141 GIDFI 145

RESULT 15  
US-09-746-660A-96  
; Sequence 96, Application US/09746660A  
; Publication No. US20030049804A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Haberhauer, Gregor  
; APPLICANT: Kim, Jun-Won  
; APPLICANT: Lee, Heung-Schick  
; APPLICANT: Hwang, Byung-Joon  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
; FILE REFERENCE: BGI-121CP2  
; CURRENT APPLICATION NUMBER: US/09/746,660A  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 09/606740  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 09/603124  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: 60/142101  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: 60/148613  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/187970  
; PRIOR FILING DATE: 2000-03-09  
; PRIOR APPLICATION NUMBER: DE 19931420.9  
; PRIOR FILING DATE: 1999-07-08  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: PatentIn Vers. 2.0  
; SEQ ID NO 96  
; LENGTH: 166  
; TYPE: PRT

; ORGANISM: Corynebacterium glutamicum  
US-09-746-660A-96

Query Match 31.2%; Score 5; DB 9; Length 166;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|  
Db 141 GIDFI 145

## RESULT 16

US-09-791-279-200  
; Sequence 200, Application US/09791279  
; Publication No. US20030050456A1  
; GENERAL INFORMATION:  
; APPLICANT: Vogel, Gabriel  
; APPLICANT: Wood, Linda S.  
; APPLICANT: Parodi, Luis  
; APPLICANT: Lind, Peter

; TITLE OF INVENTION: No. US20030050456A1e1 G Protein-Coupled Receptors  
; FILE REFERENCE: 00048.US1  
; CURRENT APPLICATION NUMBER: US/09/791,279  
; CURRENT FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: 60/184,715  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/184725  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/184,712  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/184,606  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/184,602  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/184,604  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/184,822  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/184,710  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/184,689  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/184,690  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/184,716  
; PRIOR FILING DATE: 2000-02-24  
; NUMBER OF SEQ ID NOS: 220  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 200  
; LENGTH: 198  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-791-279-200

Query Match 31.2%; Score 5; DB 9; Length 198;  
Best Local Similarity 100.0%; Pred. No. 11e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|  
Db 119 FIIFW 123

## RESULT 17

US-09-728-721-65  
; Sequence 65, Application US/09728721  
; Patent No. US20020061845A1  
; GENERAL INFORMATION:  
; APPLICANT: Bertin, John  
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES THERE  
; FILE REFERENCE: 07334-124001  
; CURRENT APPLICATION NUMBER: US/09/728,721

; CURRENT FILING DATE: 2000-12-01  
; PRIOR APPLICATION NUMBER: 09/340,620  
; PRIOR FILING DATE: 1999-06-28  
; PRIOR APPLICATION NUMBER: US 09/207,359  
; PRIOR FILING DATE: 1998-12-08  
; PRIOR APPLICATION NUMBER: US 09/099,041  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: US 09/019,942  
; PRIOR FILING DATE: 1998-02-06  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 65  
; LENGTH: 208  
; TYPE: PRT  
; ORGANISM: Rattus rattus  
US-09-728-721-65

Query Match 31.2%; Score 5; DB 10; Length 208;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FQANC 5  
|  
Db 64 FQANC 68

## RESULT 18

US-09-747-155-47  
; Sequence 47, Application US/09747155  
; Patent No. US20020151692A1  
; GENERAL INFORMATION:  
; APPLICANT: Rouquier, Sylvie  
; APPLICANT: Giorgi, Dominique  
; TITLE OF INVENTION: No. US20020151692A1e1 Polypeptides and Nucleic Acids Encod  
; FILE REFERENCE: 19904-008 (C009B6834US)  
; CURRENT APPLICATION NUMBER: US/09/747,155  
; CURRENT FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/171,746  
; PRIOR FILING DATE: 1999-12-22  
; NUMBER OF SEQ ID NOS: 431  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 47  
; LENGTH: 216  
; TYPE: PRT  
; ORGANISM: Gorilla gorilla  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(649)  
; OTHER INFORMATION: Taxon = 9593; gene = GGO19; Accession DDBJ/EMBL/GenBank =  
US-09-747-155-47

Query Match 31.2%; Score 5; DB 10; Length 216;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|  
Db 83 FIIFW 87

## RESULT 19

US-09-747-155-51  
; Sequence 51, Application US/09747155  
; Patent No. US20020151692A1  
; GENERAL INFORMATION:  
; APPLICANT: Rouquier, Sylvie  
; APPLICANT: Giorgi, Dominique  
; TITLE OF INVENTION: No. US20020151692A1e1 Polypeptides and Nucleic Acids Encod  
; FILE REFERENCE: 19904-008 (C009B6834US)  
; CURRENT APPLICATION NUMBER: US/09/747,155  
; CURRENT FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/171,746  
; PRIOR FILING DATE: 1999-12-22

; NUMBER OF SEQ ID NOS: 431  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 51  
; LENGTH: 216  
; TYPE: PRT  
; ORGANISM: Gorilla gorilla  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(649)  
; OTHER INFORMATION: Taxon = 9593; gene = GGO3; Accession DDBJ/EMBL/GenBank = AF127848  
US-09-747-155-51

Query Match 31.2%; Score 5; DB 10; Length 216;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|  
|  
|  
|  
Db 83 FIIFW 87

## RESULT 20

US-09-747-155-116  
; Sequence 116, Application US/09747155  
; Patent No. US20020151692A1  
; GENERAL INFORMATION:  
; APPLICANT: Rouquier, Sylvie  
; APPLICANT: Giorgi, Dominique  
; TITLE OF INVENTION: No. US20020151692A1 Polypeptides and Nucleic Acids Encoding Sam  
; FILE REFERENCE: 19904-008 (C009B6834US)  
; CURRENT APPLICATION NUMBER: US/09/747,155  
; CURRENT FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/171,746  
; PRIOR FILING DATE: 1999-12-22  
; NUMBER OF SEQ ID NOS: 431  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 116  
; LENGTH: 216  
; TYPE: PRT  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(649)  
; OTHER INFORMATION: Taxon = 9600; gene = PPY49; Accession DDBJ/EMBL/GenBank = AF12788

US-09-747-155-116

Query Match 31.2%; Score 5; DB 10; Length 216;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|  
|  
|  
|  
Db 83 FIIFW 87

## RESULT 21

US-10-107-868-25  
; Sequence 25, Application US/10107868  
; Patent No. US20020156242A1  
; GENERAL INFORMATION:  
; APPLICANT: Tamatani, Takuya  
; APPLICANT: Tezuka, Katsunari  
; TITLE OF INVENTION: CELL SURFACE MOLECULE MEDIATING CELL  
; TITLE OF INVENTION: ADHESION AND SIGNAL TRANSMISSION  
; FILE REFERENCE: 06501-039002  
; CURRENT APPLICATION NUMBER: US/10/107,868  
; CURRENT FILING DATE: 2002-03-26  
; PRIOR APPLICATION NUMBER: 09/561,308  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 09/383,551  
; PRIOR FILING DATE: 1999-08-26  
; PRIOR APPLICATION NUMBER: PCT/JP98/00837  
; PRIOR FILING DATE: 1998-02-27

; PRIOR APPLICATION NUMBER: JAPAN 09-62290  
; PRIOR FILING DATE: 1997-02-27  
; PRIOR APPLICATION NUMBER: JAPAN 10-62217  
; PRIOR FILING DATE: 1998-02-26  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 25  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-107-868-25

Query Match 31.2%; Score 5; DB 9; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|  
|  
|  
|  
Db 174 FIIFW 178

## RESULT 22

US-09-989-545-19  
; Sequence 19, Application US/09989545  
; Patent No. US20020164697A1  
; GENERAL INFORMATION:  
; APPLICANT: Lehar, Sophie  
; APPLICANT: Manning, Stephen  
; APPLICANT: Coyle, Anthony J.  
; APPLICANT: Gutierrez-Ramos, Jose-Carlos  
; TITLE OF INVENTION: No. US20020164697A1 Th2-Specific Molecules and Uses The  
; FILE REFERENCE: 5800-10B  
; CURRENT APPLICATION NUMBER: US/09/989,545  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 09/168,229  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 09/258,670  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-989-545-19

Query Match 31.2%; Score 5; DB 9; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|  
|  
|  
|  
Db 174 FIIFW 178

## RESULT 23

US-10-301-056-25  
; Sequence 25, Application US/10301056  
; Publication No. US20030083472A1  
; GENERAL INFORMATION:  
; APPLICANT: Tamatani, Takuya  
; APPLICANT: Tezuka, Katsunari  
; TITLE OF INVENTION: CELL SURFACE MOLECULE MEDIATING CELL  
; TITLE OF INVENTION: ADHESION AND SIGNAL TRANSMISSION  
; FILE REFERENCE: 06501-039001  
; CURRENT APPLICATION NUMBER: US/10/301,056  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US/09/383,551  
; PRIOR FILING DATE: 1999-08-26  
; PRIOR APPLICATION NUMBER: PCT/JP98/00837  
; PRIOR FILING DATE: 1998-02-27  
; PRIOR APPLICATION NUMBER: JAPAN 09-62290  
; PRIOR FILING DATE: 1997-02-27



; PRIOR APPLICATION NUMBER: JAPAN 10-62217  
; PRIOR FILING DATE: 1998-02-26  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 25  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-301-056-25

Query Match 31.2%; Score 5; DB 9; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|  
Db 174 FIIFW 178

## RESULT 24

US-10-107-828-25  
; Sequence 25, Application US/10107828  
; Patent No. US20020115831A1  
; GENERAL INFORMATION:  
; APPLICANT: Tamatani, Takuya  
; APPLICANT: Tezuka, Katsunari  
; TITLE OF INVENTION: CELL SURFACE MOLECULE MEDIATING CELL  
; TITLE OF INVENTION: ADHESION AND SIGNAL TRANSMISSION  
; FILE REFERENCE: 06501-039002  
; CURRENT APPLICATION NUMBER: US/10/107,828  
; CURRENT FILING DATE: 2002-03-26  
; PRIOR APPLICATION NUMBER: US/09/561,308B  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: PCT/JP98/00837  
; PRIOR FILING DATE: 1998-02-27  
; PRIOR APPLICATION NUMBER: JAPAN 09-62290  
; PRIOR FILING DATE: 1997-02-27  
; PRIOR APPLICATION NUMBER: JAPAN 10-62217  
; PRIOR FILING DATE: 1998-02-26  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 25  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-107-828-25

Query Match 31.2%; Score 5; DB 12; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|  
Db 174 FIIFW 178

## RESULT 25

US-10-107-907-25  
; Sequence 25, Application US/10107907  
; Patent No. US20020151685A1  
; GENERAL INFORMATION:  
; APPLICANT: Tamatani, Takuya  
; APPLICANT: Tezuka, Katsunari  
; TITLE OF INVENTION: CELL SURFACE MOLECULE MEDIATING CELL  
; TITLE OF INVENTION: ADHESION AND SIGNAL TRANSMISSION  
; FILE REFERENCE: 06501-039002  
; CURRENT APPLICATION NUMBER: US/10/107,907  
; CURRENT FILING DATE: 2002-03-26  
; PRIOR APPLICATION NUMBER: 09/561,308  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: PCT/JP98/00837  
; PRIOR FILING DATE: 1998-02-27  
; PRIOR APPLICATION NUMBER: JAPAN 09-62290

; PRIOR FILING DATE: 1997-02-27  
; PRIOR APPLICATION NUMBER: JAPAN 10-62217  
; PRIOR FILING DATE: 1998-02-26  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 25  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-107-907-25

Query Match 31.2%; Score 5; DB 12; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|  
Db 174 FIIFW 178

## RESULT 26

US-09-738-626-4341  
; Sequence 4341, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4341  
; LENGTH: 228  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4341

Query Match 31.2%; Score 5; DB 9; Length 228;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 IFWIF 15  
|  
Db 23 IFWIF 27

## RESULT 27

US-09-738-626-3808  
; Sequence 3808, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 3808  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3808

Query Match 31.2%; Score 5; DB 9; Length 305;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
Db 254 NCGID 258

## RESULT 28

US-09-864-029-2  
; Sequence 2, Application US/09864029  
; Publication No. US20030082174A1

## GENERAL INFORMATION:

; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Grosse, William M.  
; APPLICANT: Szekeres Jr., Edward S.  
; APPLICANT: Alsobrook II, John P.  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Taupier Jr., Raymond J.  
; APPLICANT: Gangolli, Esha A.  
; APPLICANT: MacDougall, John R.  
; APPLICANT: Stone, David J.  
; APPLICANT: Smithson, Glenda

; TITLE OF INVENTION: NOVEL GPCR-LIKE PROTEINS AND NUCLEIC ACIDS ENCODING  
; TITLE OF INVENTION: SAME

; FILE REFERENCE: 21402-022

; CURRENT APPLICATION NUMBER: US/09/864,029

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: 60/206,757

; PRIOR FILING DATE: 2000-05-24

; PRIOR APPLICATION NUMBER: 60/214,372

; PRIOR FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: 60/219,786

; PRIOR FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 60/207,020

; PRIOR FILING DATE: 2000-05-25

; PRIOR APPLICATION NUMBER: 60/220,593

; PRIOR FILING DATE: 2000-07-25

; PRIOR APPLICATION NUMBER: 60/239,542

; PRIOR FILING DATE: 2000-10-10

; PRIOR APPLICATION NUMBER: 60/256,402

; PRIOR FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: 60/271,645

; PRIOR FILING DATE: 2001-02-26

; PRIOR APPLICATION NUMBER: 60/274,809

; PRIOR FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: 60/275,590  
; PRIOR FILING DATE: 2001-03-13  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 312  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-864-029-2

Query Match 31.2%; Score 5; DB 9; Length 312;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIFW 13  
Db 150 FIFW 154

## RESULT 29

US-09-864-029-4

; Sequence 4, Application US/09864029

; Publication No. US20030082174A1

## GENERAL INFORMATION:

; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Grosse, William M.  
; APPLICANT: Szekeres Jr., Edward S.  
; APPLICANT: Alsobrook II, John P.  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Taupier Jr., Raymond J.  
; APPLICANT: Casman, Stacie J.  
; APPLICANT: Gangolli, Esha A.  
; APPLICANT: MacDougall, John R.  
; APPLICANT: Stone, David J.  
; APPLICANT: Smithson, Glenda

; TITLE OF INVENTION: NOVEL GPCR-LIKE PROTEINS AND NUCLEIC ACIDS ENCODING  
; TITLE OF INVENTION: SAME

; FILE REFERENCE: 21402-022

; CURRENT APPLICATION NUMBER: US/09/864,029

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: 60/206,757

; PRIOR FILING DATE: 2000-05-24

; PRIOR APPLICATION NUMBER: 60/214,372

; PRIOR FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: 60/219,786

; PRIOR FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 60/207,020

; PRIOR FILING DATE: 2000-05-25

; PRIOR APPLICATION NUMBER: 60/220,593

; PRIOR FILING DATE: 2000-07-25

; PRIOR APPLICATION NUMBER: 60/239,542

; PRIOR FILING DATE: 2000-10-10

; PRIOR APPLICATION NUMBER: 60/256,402

; PRIOR FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: 60/271,645

; PRIOR FILING DATE: 2001-02-26

; PRIOR APPLICATION NUMBER: 60/274,809

; PRIOR FILING DATE: 2001-03-09

; PRIOR APPLICATION NUMBER: 60/275,590

; PRIOR FILING DATE: 2001-03-13

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 312

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-864-029-4

Query Match 31.2%; Score 5; DB 9; Length 312;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
| | | | |  
Db 150 FIIFW 154

RESULT 30  
US-09-815-242-4990  
; Sequence 4990, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 4990  
; LENGTH: 321  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-815-242-4990

Query Match 31.2%; Score 5; DB 10; Length 321;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
| | | | |  
Db 157 FIIFW 161

RESULT 31  
US-09-815-242-10635  
; Sequence 10635, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 10635  
; LENGTH: 321  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-815-242-10635

Query Match 31.2%; Score 5; DB 10; Length 321;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
| | | | |  
Db 157 FIIFW 161

RESULT 32  
US-09-816-028A-48  
; Sequence 48, Application US/09816028A  
; Patent No. US20020042369A1  
; GENERAL INFORMATION:  
; APPLICANT: Gilbert, Michel  
; APPLICANT: Wakarchuk, Warren W.  
; APPLICANT: National Research Council of Canada  
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of  
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics  
; FILE REFERENCE: 019633-000111US  
; CURRENT APPLICATION NUMBER: US/09/816,028A  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/118,213  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: US 09/495,406  
; PRIOR FILING DATE: 2000-01-31  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 48  
; LENGTH: 322  
; TYPE: PRT  
; ORGANISM: Campylobacter jejuni  
; FEATURE:  
; OTHER INFORMATION: Campylobacter alpha-2,3-sialyltransferase I (Cst-I)  
; OTHER INFORMATION: from C. jejuni OH4384  
US-09-816-028A-48

Query Match 31.2%; Score 5; DB 10; Length 322;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGIDF 9  
| | | | |  
Db 166 CGIDF 170

RESULT 33  
US-09-864-029-8

```
; Sequence 8, Application US/09864029
; Publication No. US20030082174A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Grosse, William M.
; APPLICANT: Szekeres Jr., Edward S.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Taupier Jr., Raymond J.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: MacDougall, John R.
; APPLICANT: Stone, David J.
; APPLICANT: Smithson, Glennnda
; TITLE OF INVENTION: NOVEL GPCR-LIKE PROTEINS AND NUCLEIC ACIDS ENCODING
; FILE OF INVENTION: SAME
; FILE REFERENCE: 21402-022
; CURRENT APPLICATION NUMBER: US/09/864,029
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/206,757
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/214,372
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/219,786
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/207,020
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/220,593
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: 60/239,542
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/256,402
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/271,645
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 60/274,809
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/275,590
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-029-8
```

```
Query Match          31.2%; Score 5; DB 9; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 DFIIF 12
        |||||
Db       9 DFIIF 13
```

```
RESULT 34
US-09-886-055-293
; Sequence 293, Application US/09886055
; Patent No. US20020132273A1
; GENERAL INFORMATION:
; APPLICANT: STRYER, LUBERT
; APPLICANT: ZOZULYA, SERGEY
; TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND
; FILE REFERENCE: 078003-0277150
; CURRENT APPLICATION NUMBER: US/09/886,055
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 60/213,812
```

```
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 522
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 293
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-886-055-293
```

```
Query Match          31.2%; Score 5; DB 10; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      8 DFIIF 12
        |||||
Db       9 DFIIF 13
```

```
RESULT 35
US-09-880-137-5
; Sequence 5, Application US/09880137
; Patent No. US20020031295A1
; GENERAL INFORMATION:
; APPLICANT: Berstein, Gabriel
; TITLE OF INVENTION: METHODS OF ASSAYING FOR G
; FILE REFERENCE: MNI-131
; CURRENT APPLICATION NUMBER: US/09/880,137
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/186,706
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-137-5
```

```
Query Match          31.2%; Score 5; DB 10; Length 388;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      5 CGIDF 9
        |||||
Db      137 CGIDF 141
```

```
RESULT 36
US-09-880-137-6
; Sequence 6, Application US/09880137
; Patent No. US20020031295A1
; GENERAL INFORMATION:
; APPLICANT: Berstein, Gabriel
; TITLE OF INVENTION: METHODS OF ASSAYING FOR G
; FILE REFERENCE: MNI-131
; CURRENT APPLICATION NUMBER: US/09/880,137
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/186,706
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-137-6
```

```
Query Match          31.2%; Score 5; DB 10; Length 388;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



Tue May 13 12:12:55 2003

QY 5 CGIDE 9  
| | | | |  
Db 137 CGIDE 141

RESULT 37  
US-09-738-626-5328  
; Sequence 5328, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 5328  
; LENGTH: 403  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-5328

Query Match 31.2%; Score 5; DB 9; Length 403;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
| | | | |  
Db 32 GIDFI 36

RESULT 38  
US-09-852-053-4  
; Sequence 4, Application US/09852053  
; Patent No. US20020055141A1  
; GENERAL INFORMATION:  
; APPLICANT: BERENS, STEPHAN  
; APPLICANT: KALINOWSKI, JORN  
; APPLICANT: PUHLER, ALFRED  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM STRAIN WITH  
; TITLE OF INVENTION: ENHANCED SECRETION ACTIVITY  
; FILE REFERENCE: MAS/21123/280248  
; CURRENT APPLICATION NUMBER: US/09/852,053  
; CURRENT FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: EPO 00110021.3  
; PRIOR FILING DATE: 2000-05-12  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 403  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
; FEATURE:  
; NAME/KEY: PROPEP  
; LOCATION: (1)..(403)  
; OTHER INFORMATION: secf  
US-09-852-053-4

Query Match 31.2%; Score 5; DB 10; Length 403;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
| | | | |  
Db 32 GIDFI 36

RESULT 39  
US-09-815-242-5128  
; Sequence 5128, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5128  
; LENGTH: 411  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-5128

Query Match 31.2%; Score 5; DB 10; Length 411;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
| | | | |  
Db 304 FIIFW 308

RESULT 40  
US-10-058-636-2  
; Sequence 2, Application US/10058636  
; Publication No. US20030049270A1  
; GENERAL INFORMATION:  
; APPLICANT: Gilbert, Michel  
; APPLICANT: Wakarchuk, Warren W.  
; APPLICANT: National Research Council of Canada  
; TITLE OF INVENTION: Lipopolysaccharide alpha-2,3 Sialyltransferase of  
; TITLE OF INVENTION: Campylobacter jejuni and its Uses  
; FILE REFERENCE: 014137-013210US  
; CURRENT APPLICATION NUMBER: US/10/058,636  
; CURRENT FILING DATE: 2002-01-29  
; PRIOR APPLICATION NUMBER: US/09/272,960

; PRIOR FILING DATE: 1999-03-18  
; PRIOR APPLICATION NUMBER: US 60/078,891  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: US 09/272,960  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 430  
; TYPE: PRT  
; ORGANISM: Campylobacter jejuni  
US-10-058-636-2

Query Match 31.2%; Score 5; DB 9; Length 430;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGIDF 9  
|||||  
Db 166 CGIDF 170

## RESULT 41

US-09-992-598-177  
; Sequence 177, Application US/09992598  
; Patent No. US20020160384A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Napier, Mary A.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: P2730PLC20  
; CURRENT APPLICATION NUMBER: US/09/992,598  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/049787  
; PRIOR FILING DATE: 1997-06-16  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/065186  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/065311  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/066770  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/075945  
; PRIOR FILING DATE: 1998-02-25  
; PRIOR APPLICATION NUMBER: 60/078910  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/084600

; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/087106  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087607  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087609  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087759  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087827  
; PRIOR FILING DATE: 1998-06-03  
; PRIOR APPLICATION NUMBER: 60/088021  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088025  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088026  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088028  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088029  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088030  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088033  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088326  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088167  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088202  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088212  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088217  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088655  
; PRIOR FILING DATE: 1998-06-09  
; PRIOR APPLICATION NUMBER: 60/088734  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088738  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088742  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088810  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088824  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088826  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088858  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088861  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088876  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/089105  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089440  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089512  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089514  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089532  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089538  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089598  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089599  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089600

;; PRIOR APPLICATION NUMBER: 60/089653  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089801  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089907  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089908  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089947  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/089948  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/089952  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/090246  
;; PRIOR FILING DATE: 1998-06-22  
;; PRIOR APPLICATION NUMBER: 60/090252  
;; PRIOR FILING DATE: 1998-06-22  
;; PRIOR APPLICATION NUMBER: 60/090254  
;; PRIOR FILING DATE: 1998-06-22  
;; PRIOR APPLICATION NUMBER: 60/090349  
;; PRIOR FILING DATE: 1998-06-23  
;; PRIOR APPLICATION NUMBER: 60/090355  
;; PRIOR FILING DATE: 1998-06-23  
;; PRIOR APPLICATION NUMBER: 60/090429  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090431  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090435  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090444  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090445  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090472  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090535  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090540  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090542  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090557  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090676  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090678  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090690  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090694  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090695  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090696  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090862  
;; PRIOR FILING DATE: 1998-06-26  
;; PRIOR APPLICATION NUMBER: 60/090863  
;; PRIOR FILING DATE: 1998-06-26  
;; PRIOR APPLICATION NUMBER: 60/091360  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091478  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091544  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091519  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091626  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091633  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091978

;; PRIOR FILING DATE: 1998-07-07  
;; PRIOR APPLICATION NUMBER: 60/091982  
;; PRIOR FILING DATE: 1998-07-07  
;; PRIOR APPLICATION NUMBER: 60/092182  
;; PRIOR FILING DATE: 1998-07-09  
  
Query Match 31.2%; Score 5; DB 9; Length 445;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 8 DFIF 12  
      |||||  
Db 325 DFIF 329  
  
RESULT 42  
US-09-989-293A-177  
; Sequence 177, Application US/09989293A  
; Patent No. US20020177164A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Napier, Mary A.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: P2730PLC66  
; CURRENT APPLICATION NUMBER: US/09/989,293A  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/049787  
; PRIOR FILING DATE: 1997-06-16  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/065186  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/065311  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/066770  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/075945  
; PRIOR FILING DATE: 1998-02-25  
; PRIOR APPLICATION NUMBER: 60/078910  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/084600  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/087106  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087607  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087609  
; PRIOR FILING DATE: 1998-06-02

;	PRIOR FILING DATE:	1998-06-18
;	PRIOR APPLICATION NUMBER:	60/089947
;	PRIOR FILING DATE:	1998-06-19
;	PRIOR APPLICATION NUMBER:	60/089948
;	PRIOR FILING DATE:	1998-06-19
;	PRIOR APPLICATION NUMBER:	60/089952
;	PRIOR FILING DATE:	1998-06-19
;	PRIOR APPLICATION NUMBER:	60/090246
;	PRIOR FILING DATE:	1998-06-22
;	PRIOR APPLICATION NUMBER:	60/090252
;	PRIOR FILING DATE:	1998-06-22
;	PRIOR APPLICATION NUMBER:	60/090254
;	PRIOR FILING DATE:	1998-06-22
;	PRIOR APPLICATION NUMBER:	60/090349
;	PRIOR FILING DATE:	1998-06-23
;	PRIOR APPLICATION NUMBER:	60/090355
;	PRIOR FILING DATE:	1998-06-23
;	PRIOR APPLICATION NUMBER:	60/090429
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090431
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090435
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090444
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090445
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090472
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090535
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090540
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090542
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090557
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090676
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090678
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090690
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090694
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090695
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090696
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090862
;	PRIOR FILING DATE:	1998-06-26
;	PRIOR APPLICATION NUMBER:	60/090863
;	PRIOR FILING DATE:	1998-06-26
;	PRIOR APPLICATION NUMBER:	60/091360
;	PRIOR FILING DATE:	1998-07-01
;	PRIOR APPLICATION NUMBER:	60/091478
;	PRIOR FILING DATE:	1998-07-02
;	PRIOR APPLICATION NUMBER:	60/091544
;	PRIOR FILING DATE:	1998-07-01
;	PRIOR APPLICATION NUMBER:	60/091519
;	PRIOR FILING DATE:	1998-07-02
;	PRIOR APPLICATION NUMBER:	60/091626
;	PRIOR FILING DATE:	1998-07-02
;	PRIOR APPLICATION NUMBER:	60/091633
;	PRIOR FILING DATE:	1998-07-02
;	PRIOR APPLICATION NUMBER:	60/091978
;	PRIOR FILING DATE:	1998-07-07
;	PRIOR APPLICATION NUMBER:	60/091982
;	PRIOR FILING DATE:	1998-07-07
;	PRIOR APPLICATION NUMBER:	60/092182
;	PRIOR FILING DATE:	1998-07-09

31.28; Score 5; DB 9; Length 445;



Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
|||||  
Db 325 DFIIF 329

RESULT 43

US-10-063-547-32  
; Sequence 32, Application US/10063547  
; Publication No. US20020182638A1  
; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/063,547  
; CURRENT FILING DATE: 2002-05-02  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 32  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-063-547-32

Query Match 31.2%; Score 5; DB 9; Length 445;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
|||||  
Db 325 DFIIF 329

RESULT 44

US-09-989-735-177  
; Sequence 177, Application US/09989735  
; Publication No. US20020193299A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Napier, Mary A.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: P2730P1C61  
; CURRENT APPLICATION NUMBER: US/09/989,735  
; CURRENT FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: 60/049787  
; PRIOR FILING DATE: 1997-06-16  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/065186  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/065311  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/066770  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/075945  
; PRIOR FILING DATE: 1998-02-25  
; PRIOR APPLICATION NUMBER: 60/078910  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/084600  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/087106  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087607  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087609  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087759  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087827  
; PRIOR FILING DATE: 1998-06-03  
; PRIOR APPLICATION NUMBER: 60/088021  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088025  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088026  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088028  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088029  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088030  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088033  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088326  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088167  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088202  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088212  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088217  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088655  
; PRIOR FILING DATE: 1998-06-09  
; PRIOR APPLICATION NUMBER: 60/088734  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088738  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088742  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088810  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088824  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088826  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088858  
; PRIOR FILING DATE: 1998-06-11

;  
; PRIOR APPLICATION NUMBER: 60/088861  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088876  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/089105  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089440  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089512  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089514  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089532  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089538  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089598  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089599  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089600  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089653  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089801  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/089907  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/089908  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/089947  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: 60/089948  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: 60/089952  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: 60/090246  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090252  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090254  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090349  
; PRIOR FILING DATE: 1998-06-23  
; PRIOR APPLICATION NUMBER: 60/090355  
; PRIOR FILING DATE: 1998-06-23  
; PRIOR APPLICATION NUMBER: 60/090429  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090431  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090435  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090444  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090445  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090472  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090535  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090540  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090542  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090557  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090676  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090678  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090690  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090694

;  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090695  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090696  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090862  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: 60/090863  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: 60/091360  
; PRIOR FILING DATE: 1998-07-01  
; PRIOR APPLICATION NUMBER: 60/091478  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091544  
; PRIOR FILING DATE: 1998-07-01  
; PRIOR APPLICATION NUMBER: 60/091519  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091626  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091633  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091978  
; PRIOR FILING DATE: 1998-07-07  
; PRIOR APPLICATION NUMBER: 60/091982  
; PRIOR FILING DATE: 1998-07-07  
; PRIOR APPLICATION NUMBER: 60/092182  
; PRIOR FILING DATE: 1998-07-09

Query Match 31.2% Score 5; DB 9; Length 445;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12  
|||||  
Db 325 DFIIIF 329

RESULT 45

US-09-990-444-177  
; Sequence 177, Application US/09990444  
; Publication No. US20020193300A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Napier, Mary A.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: P2730P1C19  
; CURRENT APPLICATION NUMBER: US/09/990,444  
; CURRENT FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/049787  
; PRIOR FILING DATE: 1997-06-16

;	PRIOR FILING DATE:	1998-06-16
;	PRIOR APPLICATION NUMBER:	60/089512
;	PRIOR FILING DATE:	1998-06-16
;	PRIOR APPLICATION NUMBER:	60/089514
;	PRIOR FILING DATE:	1998-06-16
;	PRIOR APPLICATION NUMBER:	60/089532
;	PRIOR FILING DATE:	1998-06-17
;	PRIOR APPLICATION NUMBER:	60/089538
;	PRIOR FILING DATE:	1998-06-17
;	PRIOR APPLICATION NUMBER:	60/089598
;	PRIOR FILING DATE:	1998-06-17
;	PRIOR APPLICATION NUMBER:	60/089599
;	PRIOR FILING DATE:	1998-06-17
;	PRIOR APPLICATION NUMBER:	60/089600
;	PRIOR FILING DATE:	1998-06-17
;	PRIOR APPLICATION NUMBER:	60/089653
;	PRIOR FILING DATE:	1998-06-17
;	PRIOR APPLICATION NUMBER:	60/089801
;	PRIOR FILING DATE:	1998-06-18
;	PRIOR APPLICATION NUMBER:	60/089907
;	PRIOR FILING DATE:	1998-06-18
;	PRIOR APPLICATION NUMBER:	60/089908
;	PRIOR FILING DATE:	1998-06-18
;	PRIOR APPLICATION NUMBER:	60/089947
;	PRIOR FILING DATE:	1998-06-19
;	PRIOR APPLICATION NUMBER:	60/089948
;	PRIOR FILING DATE:	1998-06-19
;	PRIOR APPLICATION NUMBER:	60/089952
;	PRIOR FILING DATE:	1998-06-19
;	PRIOR APPLICATION NUMBER:	60/090246
;	PRIOR FILING DATE:	1998-06-22
;	PRIOR APPLICATION NUMBER:	60/090252
;	PRIOR FILING DATE:	1998-06-22
;	PRIOR APPLICATION NUMBER:	60/090254
;	PRIOR FILING DATE:	1998-06-22
;	PRIOR APPLICATION NUMBER:	60/090349
;	PRIOR FILING DATE:	1998-06-23
;	PRIOR APPLICATION NUMBER:	60/090355
;	PRIOR FILING DATE:	1998-06-23
;	PRIOR APPLICATION NUMBER:	60/090429
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090431
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090435
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090444
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090445
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090472
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090535
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090540
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090542
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090557
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090676
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090678
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090690
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090694
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090695
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090696
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090862
;	PRIOR FILING DATE:	1998-06-26

; PRIOR APPLICATION NUMBER: 60/090863  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: 60/091360  
; PRIOR FILING DATE: 1998-07-01  
; PRIOR APPLICATION NUMBER: 60/091478  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091544  
; PRIOR FILING DATE: 1998-07-01  
; PRIOR APPLICATION NUMBER: 60/091519  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091626  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091633  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091978  
; PRIOR FILING DATE: 1998-07-07  
; PRIOR APPLICATION NUMBER: 60/091982  
; PRIOR FILING DATE: 1998-07-07  
; PRIOR APPLICATION NUMBER: 60/092182  
; PRIOR FILING DATE: 1998-07-09

Query Match 31.2%; Score 5; DB 9; Length 445;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIF 12

Db 325 DFIF 329

Search completed: May 11, 2003, 20:21:14  
Job time : 6.23894 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:07:52 ; Search time 4.95575 Seconds  
(without alignments)  
310.377 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	43.8	483	S41689	cytochrome-c oxida
2	6	37.5	156	G69233	N-terminal acetyl
3	6	37.5	205	E64621	conserved hypothet
4	6	37.5	283	E64361	nicotinate-nucleot
5	6	37.5	330	T25939	hypothetical prote
6	6	37.5	360	T35783	probable secreted
7	6	37.5	372	JQ2135	NADH2 dehydrogenas
8	6	37.5	430	T46216	hypothetical prote
9	6	37.5	590	1 QQUTC5	NADH2 dehydrogenas
10	6	37.5	679	1 QQCV5	trifunctional enzy
11	6	37.5	1091	2 F83928	hypothetical prote
12	5	31.2	41	2 G81376	hypothetical prote
13	5	31.2	106	2 S76060	hypothetical prote
14	5	31.2	120	2 D89910	large-conductance
15	5	31.2	120	2 AH1280	hypothetical prote
16	5	31.2	122	2 G97899	hypothetical prote
17	5	31.2	128	1 S27723	cytochrome cytm pr
18	5	31.2	128	2 A69214	hypothetical prote
19	5	31.2	128	2 G64155	yhcC protein - Hae
20	5	31.2	135	2 B31933	Ig heavy chain pre
21	5	31.2	136	2 T16723	hypothetical prote
22	5	31.2	138	2 H97024	uncharacterized sm
23	5	31.2	160	2 B96986	probable tRNA-meth
24	5	31.2	163	2 AF2342	hypothetical prote
25	5	31.2	173	2 B89885	conserved hypothet
26	5	31.2	180	2 B98134	hypothetical prote
27	5	31.2	180	2 AH3153	hypothetical prote
28	5	31.2	181	2 G82911	hypothetical prote
29	5	31.2	187	2 E81383	probable integral

30	5	31.2	195	2	H86768
31	5	31.2	195	2	A48957
32	5	31.2	205	2	JN0755
33	5	31.2	208	2	AC1787
34	5	31.2	208	2	AD1411
35	5	31.2	210	2	AC2982
36	5	31.2	214	2	T22896
37	5	31.2	214	2	T22892
38	5	31.2	219	2	D81792
39	5	31.2	220	1	RWHU28
40	5	31.2	220	2	B90393
41	5	31.2	225	1	C64201
42	5	31.2	225	2	AC0587
43	5	31.2	230	2	H97331
44	5	31.2	230	2	T32999
45	5	31.2	233	2	AF0367
46	5	31.2	247	2	C84913
47	5	31.2	248	2	H70550
48	5	31.2	248	2	AF2533
49	5	31.2	250	2	E70452
50	5	31.2	258	2	S26762
51	5	31.2	260	2	T10642
52	5	31.2	264	2	JC7772
53	5	31.2	265	2	E96992
54	5	31.2	268	2	G83640
55	5	31.2	268	2	F72240
56	5	31.2	269	1	JQ2127
57	5	31.2	269	1	A30768
58	5	31.2	272	2	T28737
59	5	31.2	276	2	G85074
60	5	31.2	277	2	T50805
61	5	31.2	278	2	H86824
62	5	31.2	278	2	E98301
63	5	31.2	282	2	T33036
64	5	31.2	288	1	S73464
65	5	31.2	288	2	S75324

ALIGNMENTS

RESULT 1

S41689  
cytochrome-c oxidase (EC 1.9.3.1) chain I - Theileria parva mitochondrion (fragme  
C;Species: mitochondrion Theileria parva  
C;Date: 31-Dec-1993 #sequence\_revision 02-Aug-1994 #text\_change 07-Dec-1999  
C;Accession: S41689; S40154  
R;Kairo, A.; Fairlamb, A.H.; Gobright, E.; Nene, V.  
EMBO J. 13, 898-905, 1994  
A;Title: A 7.1 kb linear DNA molecule of Theileria parva has scrambled rDNA sequ  
A;Reference number: S41689; MUID:94155854; PMID:8112303  
A;Accession: S41689  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-483 <KAI>  
A;Cross-references: EMBL:Z23263; NID:g437862; PIDN:CAA80798.1; PID:g437863  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1993  
C;Genetics:  
A;Genome: mitochondrion  
A;Genetic code: SGC6  
C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homolog  
C;Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; memb  
transmembrane protein  
F;18-466/Domain: cytochrome-c oxidase chain I homology <CO1>  
F;71,386/Binding site: heme a iron (His) (axial ligands) #status predicted  
F;248-252/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
F;252/Binding site: oxygen (Tyr) #status predicted  
F;384/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 43.8%; Score 7; DB 2; Length 483;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFIIF 12  
| | | | |  
Db 150 GIDFIIF 156

## RESULT 2

G69233  
N-terminal acetyltransferase complex, subunit ARD1 - Methanobacterium thermoautotrophicum  
C:Species: Methanobacterium thermoautotrophicum  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 08-Oct-1999  
C:Accession: G69233  
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwanli, N.; ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: function  
A:Reference number: A69000; MUID:98037514; PMID:9371463  
A:Accession: G69233  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-156 <MTH>  
A:Cross-references: GB:AE000872; GB:AE000666; NID:g2622082; PIDN:AAB85496.1; PID:g262210  
A:Experimental source: strain Delta H  
C:Genetics:  
A:Gene: MTH999  
C:Superfamily: Escherichia coli ribosomal-protein-alanine N-acetyltransferase rimi

Query Match 37.5%; Score 6; DB 2; Length 156;  
Best Local Similarity 100.0%; Pred. No. 6.2;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIIFI 14  
| | | | |  
Db 53 FIIIFI 58

## RESULT 3

E64621  
Conserved hypothetical protein HP0813 - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 24-Sep-1999  
C:Accession: E64621  
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467; PMID:9252185  
A:Accession: E64621  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-205 <TOM>  
A:Cross-references: GB:AE000593; GB:AE000511; NID:g2313944; PIDN:AAD07862.1; PID:g231394  
C:Genetics:  
A:Start codon: TTG  
C:Superfamily: glyoxalase

Query Match 37.5%; Score 6; DB 2; Length 205;  
Best Local Similarity 100.0%; Pred. No. 7.7;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11  
| | | | |  
Db 23 GIDFII 28

## RESULT 4

E64361  
nicotinate-nucleotide diphosphorylase (carboxylating) (EC 2.4.2.19) - Methanococcus jann  
N:Alternate names: quinolinate phosphoribosyltransferase  
C:Species: Methanococcus jannaschii  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 03-Jun-2002

C:Accession: E64361  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glorion, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A:Reference number: A64300; MUID:96337999; PMID:8688087  
A:Accession: E64361  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-283 <BUL>  
A:Cross-references: GB:U67499; GB:L77117; NID:g1591190; PIDN:AAB98483.1; PID:g1591190  
C:Genetics:  
A:Map position: REV436372-435521  
C:Superfamily: nicotinate-nucleotide pyrophosphorylase (carboxylating)  
C:Keywords: glycosyltransferase; pentosyltransferase

Query Match 37.5%; Score 6; DB 2; Length 283;  
Best Local Similarity 100.0%; Pred. No. 9.9;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGIDFI 10  
| | | | |  
Db 48 CGIDFI 53

## RESULT 5

T25939  
hypothetical protein ZC142.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T25939  
R:Bradshaw, H.  
submitted to the EMBL Data Library, July 1996  
A:Description: The sequence of C. elegans cosmid ZC142.  
A:Reference number: Z20114  
A:Accession: T25939  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-330 <BRA>  
A:Cross-references: EMBL:U64841; PIDN:AAB04845.1; GSPDB:GN00023; CESP:ZC142.1  
A:Experimental source: strain Bristol N2; clone ZC142  
C:Genetics:  
A:Gene: CESP:ZC142.1  
A:Map position: 5  
A:Introns: 23/3; 164/3; 213/3

Query Match 37.5%; Score 6; DB 2; Length 330;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWIF 15  
| | | | |  
Db 133 IIFWIF 138

## RESULT 6

T35783  
probable secreted protein - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T35783  
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, July 1998  
A:Reference number: Z21570  
A:Accession: T35783  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-360 <SEE>  
A:Cross-references: EMBL:AL031013; PIDN:CAAL9788.1; GSPDB:GN00070; SCOEDB:SC8A6.1  
A:Experimental source: strain A3(2)  
C:Genetics:

Query Match 37.5%; Score 6; DB 2; Length 330;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

A;Gene: SCOEDB:SC8A6.16

Query Match 37.5%; Score 6; DB 2; Length 360;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FQANCG 6

|||||

Db 90 FQANCG 95

RESULT 7

JQ2135

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - Plectonema boryanum

C;Species: Plectonema boryanum

C;Date: 28-Aug-1985 #sequence\_revision 07-Oct-1994 #text\_change 03-Jun-2002

C;Accession: JQ2135

R;Takahashi, Y.; Shonai, F.; Fujita, Y.; Kohchi, T.; Ohyama, K.; Matsubara, H.

Plant Cell Physiol. 32, 969-981, 1991

A;Title: Structure of a co-transcribed gene cluster, ndh1-frxB-ndh6-ndh4L, cloned from t

A;Reference number: JQ2135

A;Accession: JQ2135

A;Molecule type: DNA

A;Residues: 1-372 <TAK>

A;Cross-references: DDBJ:D01014; NID:g216817; PIDN:BAA00814.1; PID:g216818

A;Experimental source: strain M101

C;Genetics:

A;Gene: ndh1

C;Superfamily: NADH dehydrogenase (ubiquinone) chain 1

C;Keywords: membrane-associated complex; NAD; oxidoreductase

Query Match

Best Local Similarity 37.5%; Score 6; DB 2; Length 372;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFWI 14

|||||

Db 216 FIIFWI 221

RESULT 8

T46216

hypothetical protein T8P19.230 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 17-Mar-2000

C;Accession: T46216

R;Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa

submitted to the Protein Sequence Database, December 1999

A;Reference number: Z23008

A;Accession: T46216

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-430 <CHO>

A;Cross-references: EMBL:AL133315

A;Experimental source: cultivar Columbia; BAC clone T8P19

C;Genetics:

A;Map position: 3

A;Introns: 144/3

A;Note: T8P19.230

C;Superfamily: Arabidopsis CER2 protein

Query Match

Best Local Similarity 37.5%; Score 6; DB 2; Length 430;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ANCGID 8

|||||

Db 103 ANCGID 108

RESULT 9

QQUTC5

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Trypanosoma brucei mitochondrion

C;Species: mitochondrion Trypanosoma brucei

C;Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 03-Jun-2002

C;Accession: A04519; D22845

R;Hensgens, L.A.M.; Brakenhoff, J.; De Vries, B.F.; Sloof, P.; Tromp, M.C.; Van B

Nucleic Acids Res. 12, 7327-7344, 1984

A;Title: The sequence of the gene for cytochrome c oxidase subunit I, a frameshift

ondrial maxi-circle DNA.

A;Reference number: A93537; MUID:85037915; PMID:6093040

A;Accession: A04519

A;Molecule type: DNA

A;Residues: 1-590 <HEN>

A;Cross-references: GB:M94286; NID:g343546

A;Note: this translation is not annotated in GenBank entry TRBKPGEN, release 109.

C;Comment: The DNA sequence is from a segment of the 20-kb maxicircle, which is b

C;Genetics:

A;Genome: mitochondrion

A;Genetic code: SGC6

C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5

C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphoryl

Query Match

Best Local Similarity 37.5%; Score 6; DB 1; Length 590;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWIF 15

|||||

Db 115 IIFWIF 120

RESULT 10

QQCV5

trifunctional enzyme - cauliflower mosaic virus

N;Alternate names: ORF5 protein

N;Contains: aspartic proteinase (EC 3.4.23.-); endonuclease; RNA-directed DNA pol

C;Species: cauliflower mosaic virus

C;Date: 31-Oct-1980 #sequence\_revision 31-Oct-1980 #text\_change 23-Jul-1999

C;Accession: D90799; D93729; D94613; JN0497; A04159

R;Franck, A.; Guillely, H.; Jonard, G.; Richards, K.; Hirth, L.

Cell 21, 285-294, 1980

A;Title: Nucleotide sequence of cauliflower mosaic virus DNA.

A;Reference number: A90799; MUID:81001865; PMID:7407912

A;Accession: D90799

A;Molecule type: DNA

A;Residues: 1-679 <FRA>

A;Cross-references: GB:J02048; NID:g58821; PIDN:CAA23460.1; PID:g58821

A;Experimental source: strain Strasbourg

R;Gardner, R.C.; Howarth, A.J.; Hahn, P.; Brown-Luedi, M.; Shepherd, R.J.; Messin

Nucleic Acids Res. 9, 2871-2888, 1981

A;Title: The complete nucleotide sequence of an infectious clone of cauliflower m

A;Reference number: A93729; MUID:82014878; PMID:6269062

A;Accession: D93729

A;Molecule type: DNA

A;Residues: 1,'N',3-11,'I',13-98,'K',100-144,'T',146-154,'I',156-313,'I',315-514,

A;Experimental source: strain CM1841

R;Guillely, H.

submitted to the Nucleic Acid Sequence Database, October 1982

A;Reference number: A94613

A;Accession: D94613

A;Molecule type: DNA

A;Residues: 1-5,'Q',7-9,'I',11,'NQ',12-17,'I',19-63,'I',65-82,'N',84-86,'R',88-91

A;Experimental source: strain D/H

R;Chenault, K.D.; Melcher, U.

Gene 123, 255-257, 1993

A;Title: The complete nucleotide sequence of cauliflower mosaic virus isolate BB

A;Reference number: JN0493; MUID:93154593; PMID:8428667

A;Accession: JN0497

A;Molecule type: DNA

A;Residues: 1-56,'L',58-94,'R',96-98,'K',100-185,'K',187-313,'I',315-514,'K',516

A;Experimental source: isolate BBC

A;Note: the authors translated the codon CTC for residue 57 as Val

C;Superfamily: cauliflower mosaic virus trifunctional enzyme

C;Keywords: aspartic proteinase; endonuclease; hydrolase; multifunctional enzyme

Query Match 37.5%; Score 6; DB 1; Length 679;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11  
 |||||  
 Db 109 GIDFII 114

RESULT 11  
 F83928  
 hypothetical protein BH2230 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: F83928  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: F83928  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1091 <STO>  
 A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB05949.1; GSPDB:GN00  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH2230

Query Match 37.5%; Score 6; DB 2; Length 1091;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11  
 |||||  
 Db 880 GIDFII 885

RESULT 12  
 G81376  
 hypothetical protein Cj0344 [imported] - Campylobacter jejuni (strain NCTC 11168)  
 C:Species: Campylobacter jejuni  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
 C:Accession: G81376  
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel  
 Nature 403, 665-668, 2000  
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
 A:Reference number: A81250; MUID:20150912; PMID:10688204  
 A:Accession: G81376  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-41 <PAR>  
 A:Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB74181.1; PID:g696781  
 A:Experimental source: serotype O2, strain NCTC 11168  
 C:Genetics:  
 A:Gene: Cj0344

Query Match 31.2%; Score 5; DB 2; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12  
 |||||  
 Db 10 DFIIIF 14

RESULT 13  
 S76060  
 hypothetical protein - Synechocystis sp. (strain PCC 6803)  
 C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
 C:Accession: S76060

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyaj  
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synece  
 s.  
 A:Reference number: S74322; MUID:97061201; PMID:8905231  
 A:Accession: S76060  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-106 <KAN>  
 A:Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BAA10038.1; PID  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 199

Query Match 31.2%; Score 5; DB 2; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12  
 |||||  
 Db 97 DFIIIF 101

RESULT 14  
 D89910  
 large-conductance mechanosensitive channel [imported] - Staphylococcus aureus (s  
 C:Species: Staphylococcus aureus  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C:Accession: D89910  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, I  
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu  
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: A89758; MUID:21311952; PMID:11418146  
 A:Accession: D89910  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-120 <KUR>  
 A:Cross-references: GB:BA000018; PID:g13701145; PIDN:BAB42440.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: mscl  
 C:Superfamily: yhcC protein

Query Match 31.2%; Score 5; DB 2; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
 |||||  
 Db 71 IDFII 75

RESULT 15  
 AH1280  
 hypothetical protein lmol648 [imported] - Listeria monocytogenes (strain EGD-e)  
 C:Species: Listeria monocytogenes  
 C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AH1280  
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; B  
 .; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; F  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam,  
 ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; W  
 A:Title: Comparative genomics of Listeria species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AH1280  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-120 <GLA>  
 A:Cross-references: GB:NC\_003210; PIDN:CAC99726.1; PID:g16411084; GSPDB:GN00177  
 A:Experimental source: strain EGD-e



C;Genetics:  
A;Gene: lmo1648

Query Match 31.2%; Score 5; DB 2; Length 120;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
|||||  
Db 27 IDFII 31

RESULT 16  
G97899

hypothetical protein ABC-SBP-truncation [imported] - Streptococcus pneumoniae (strain R6  
C;Species: Streptococcus pneumoniae  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
C;Accession: G97899  
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A;Reference number: A97872; MUID:21429245; PMID:11544234  
A;Accession: G97899  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-122 <KUR>  
A;Cross-references: GB:AE007317; PIDN:AAK99027.1; PID:g15457769; GSPDB:GN00174  
C;Genetics:  
A;Gene: ABC-SBP-truncation

Query Match 31.2%; Score 5; DB 2; Length 122;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
|||||  
Db 53 IDFII 57

RESULT 17  
S27723

cytochrome cytm precursor - Synechocystis sp. (strain PCC 6803)  
N;Alternate names: protein sl11245  
C;Species: Synechocystis sp.  
A;Variety: PCC 6803  
C;Date: 17-Apr-1993 #sequence\_revision 23-May-1997 #text\_change 16-Jun-2000  
C;Accession: S75611; S27723  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
S.  
A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S75611  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-128 <KAN>  
A;Cross-references: EMBL:D90912; GB:AB001339; NID:g1653228; PIDN:BAA18172.1; PID:g165325  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
R;Malakhov, M.P.; Wada, H.; Los, D.A.; Sakamoto, T.; Murata, N.  
submitted to the EMBL Data Library, April 1992  
A;Description: Structure and expression of the cytm gene, encoding cytochrome from synec  
A;Reference number: S27720  
A;Accession: S27723  
A;Molecule type: DNA  
A;Residues: 24-128 <MAL>  
A;Cross-references: EMBL:D10716; NID:g217098; PIDN:BAA01559.1; PID:g217102  
C;Genetics:  
A;Gene: cytm  
C;Superfamily: cytochrome c6; cytochrome c6 homology

C;Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein  
F;55-128/Product: cytochrome cytm #status predicted <MAT>  
F;57-128/Domain: cytochrome c6 homology #status atypical <CYC>  
F;68,71/Binding site: heme (Cys) (covalent) #status predicted  
F;72,108/Binding site: heme iron (His, Met) (axial ligands) #status predicted

Query Match 31.2%; Score 5; DB 1; Length 128;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FQANC 5  
|||||  
Db 64 FQANC 68

RESULT 18  
A69214

hypothetical protein MTH854 - Methanobacterium thermoautotrophicum (strain Delta H  
C;Species: Methanobacterium thermoautotrophicum  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999  
C;Accession: A69214  
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge  
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwa  
kl, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H:  
A;Reference number: A69000; MUID:98037514; PMID:9371463  
A;Accession: A69214  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-128 <MTH>  
A;Cross-references: GB:AE000862; GB:AE000666; NID:g2621943; PIDN:AAB85352.1; PID:g  
A;Experimental source: strain Delta H  
C;Genetics:  
A;Gene: MTH854  
C;Superfamily: conserved hypothetical protein slr2059; ferredoxin 2[4Fe-4S] homol  
F;72-127/Domain: ferredoxin 2[4Fe-4S] homology <FER>

Query Match 31.2%; Score 5; DB 2; Length 128;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|||||  
Db 54 GIDFI 58

RESULT 19  
G64155

yhdC protein - Haemophilus influenzae (strain Rd KW20)  
C;Species: Haemophilus influenzae  
C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 29-Sep-1999  
C;Accession: G64155  
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerla  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Wei  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.  
Science 269, 496-512, 1995  
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; V  
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd  
A;Reference number: A64000; MUID:95350630; PMID:7542800  
A;Accession: G64155  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-128 <TIGR>  
A;Cross-references: GB:U32745; GB:L42023; NID:g1573617; PIDN:AAC22286.1; PID:g157  
C;Genetics:  
A;Gene: yhdC  
C;Superfamily: yhdC protein

Query Match 31.2%; Score 5; DB 2; Length 128;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11  
|||||  
Db 83 IDFI 87

RESULT 20  
B31933

Ig heavy chain precursor V region (X1914) - African clawed frog (fragment)  
C:Species: Xenopus laevis (African clawed frog)  
C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 23-Jul-1999  
C:Accession: B31933

R:Schwager, J.; Mikoryak, C.A.; Steiner, L.A.  
Proc. Natl. Acad. Sci. U.S.A. 85, 2245-2249, 1988

A:Title: Amino acid sequence of heavy chain from Xenopus laevis IgM deduced from cDNA se

A:Reference number: A94192; MUID:88176921; PMID:2451244

A:Accession: B31933

A:Molecule type: mRNA

A:Residues: 1-135 <SCH>

A:Cross-references: GB:J03632; NID:G214329; PIDN:AAA49791.1; PID:G214330

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin

F:32-114/Domain: immunoglobulin homology <IMM>

Query Match 31.2%; Score 5; DB 2; Length 135;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
|||||  
Db 1 DFIIF 5

RESULT 21  
T16723

hypothetical protein R12B2.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T16723

R:Miller, N.

submitted to the EMBL Data Library, June 1994

A:Description: The sequence of C. elegans cosmid R12B2.

A:Reference number: Z18567

A:Accession: T16723

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Residues: 1-136 <MIL>

A:Cross-references: EMBL:U00066; NID:G495688; PID:G495691; PIDN:AAA50741.1; CESP:R12B2.3

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:R12B2.3

A:Introns: 25/1; 61/3; 86/2; 112/3

## Query Match

Best Local Similarity 31.2%; Score 5; DB 2; Length 136;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
|||||  
Db 25 DFIIF 29

RESULT 22  
H97024

uncharacterized small membrane protein, homolog of ykva B. subtilis [imported] - Clostri  
C:Species: Clostridium acetobutylicum  
C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: H97024

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: H97024

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-138 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK78987.1; PID:G15023920; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC1011

## Query Match

Best Local Similarity 31.2%; Score 5; DB 2; Length 138;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWI 14  
|||||  
Db 117 IIFWI 121

RESULT 23  
B96986

probable tRNA-methylase (SpoU class) [Imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum  
C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: B96986

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacter

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B96986

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-160 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK78677.1; PID:G15023579; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC0700

C:Superfamily: Chlamydomonada pneumoniae rRNA methylase

## Query Match

Best Local Similarity 31.2%; Score 5; DB 2; Length 160;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
|||||  
Db 101 DFIIF 105

RESULT 24  
AF2342

hypothetical protein alr4293 [Imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AF2342

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iri

NA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacteriu

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AF2342

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-163 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA075992.1; PID:G17133429; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr4293

Query Match 31.2%; Score 5; DB 2; Length 163;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 IFWIF 15

Db 58 IFWIF 62  
|||||

## RESULT 25

B89885  
conserved hypothetical protein SA0989 [imported] - Staphylococcus aureus (strain N315)  
C;Species: Staphylococcus aureus  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C;Accession: B89885  
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogino, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A;Reference number: A89758; MUID:21311952; PMID:11418146  
A;Accession: B89885  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-173 <KUR>  
A;Cross-references: GB:BA000018; PID:g13700942; PIDN:BAB42238.1; GSPDB:GN00149  
A;Experimental source: strain N315  
C;Genetics:  
A;Gene: SA0989

Query Match 31.2%; Score 5; DB 2; Length 173;

Best Local Similarity 100.0%; Pred. No. 92;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11

Db 3 IDFII 7

## RESULT 26

B98134  
hypothetical protein AGR\_L70 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
C;Species: Agrobacterium tumefaciens  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C;Accession: B98134  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A;Reference number: A97359; PMID:11743194  
A;Accession: B98134  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-180 <KUR>  
A;Cross-references: GB:AE007870; PIDN:AAK88596.1; PID:g15158309; GSPDB:GN00170  
C;Genetics:  
A;Gene: AGR\_L70  
A;Map position: linear chromosome

Query Match 31.2%; Score 5; DB 2; Length 180;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 FWIFW 16

Db 47 FWIFW 51

## RESULT 27

AH3153  
hypothetical protein Atu4852 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C;Accession: AH3153  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan, Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001

A;Authors: YOO, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-ster, E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; PMID:11743193  
A;Accession: AH3153  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-180 <KUR>  
A;Cross-references: GB:AE008689; PIDN:AAL45646.1; PID:g17743370; GSPDB:GN00187  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: Atu4852  
A;Map position: linear chromosome

Query Match 31.2%; Score 5; DB 2; Length 180;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 FWIFW 16

Db 47 FWIFW 51

## RESULT 28

G82911  
hypothetical protein UU275 [imported] - Ureaplasma urealyticum  
C;Species: Ureaplasma urealyticum  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C;Accession: G82911  
R;Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.; submitted to GenBank, February 2000  
A;Description: The complete sequence of Ureaplasma urealyticum: Alternate views of  
A;Reference number: A82870  
A;Accession: G82911  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-181 <GLA>  
A;Cross-references: GB:AE002124; GB:AF222894; NID:g6899241; PIDN:AAF30684.1; GSPDB:GN00124  
A;Experimental source: serovar 3; biovar 1  
C;Genetics:  
A;Gene: UU275  
A;Genetic code: SGC3

Query Match 31.2%; Score 5; DB 2; Length 181;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IFWIF 14

Db 102 IFWIF 106

## RESULT 29

E81383  
probable integral membrane protein Cj0399 [imported] - Campylobacter jejuni (strain C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C;Accession: E81383  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals  
A;Reference number: A81250; MUID:20150912; PMID:10688204  
A;Accession: E81383  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-187 <PAR>  
A;Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB74235.1; PID:AL139075  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: Cj0399

Query Match 31.2%; Score 5; DB 2; Length 187;

Best Local Similarity 100.0%; Pred. No. 98;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWI 14  
|  
|  
|  
|  
Db 77 IIFWI 81

RESULT 30  
H86768

GTP-binding protein [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C;Species: Lactococcus lactis subsp. lactis  
C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C;Accession: H86768  
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissenbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss  
A;Reference number: A86625; MUID:21235186; PMID:11337471  
A;Accession: H86768  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-195 <STO>  
A;Cross-references: GB:AE005176; PID:gl2724115; PIDN:AAK05250.1; GSPDB:GN00146  
A;Experimental source: strain IL1403  
C;Genetics:  
A;Gene: ysxL  
C;Superfamily: Bacillus subtilis conserved hypothetical protein ysxC; translation elonga

Query Match 31.2%; Score 5; DB 2; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIF 12  
|  
|  
|  
|  
Db 172 DFIF 176

RESULT 31

A48957  
orf 2 5' of pepC - Lactococcus lactis (fragment)

C;Species: Lactococcus lactis  
C;Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 15-Oct-1999  
C;Accession: A48957  
R;Chapot-Chartier, M.P.; Nardl, M.; Chopin, M.C.; Chopin, A.; Gripon, J.C.  
Appl. Environ. Microbiol. 59, 330-333, 1993  
A;Title: Cloning and sequencing of pepC, a cysteine aminopeptidase gene from Lactococcus  
A;Reference number: A48957; MUID:93175873; PMID:8439160  
A;Accession: A48957  
A;Status: preliminary  
A;Molecule type: nucleic acid  
A;Residues: 1-195 <CHA>  
A;Cross-references: GB:M86245; NID:g293010; PIDN:AAA74513.1; PID:g293011  
A;Experimental source: subsp. cremoris AM2  
A;Note: sequence extracted from NCBI backbone (NCBIN:125564, NCBIP:125565)

Query Match 31.2%; Score 5; DB 2; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
|  
|  
|  
|  
Db 168 IDFII 172

RESULT 32  
JN0755

urease accessory protein ureG - Proteus mirabilis  
N;Contains: probable urease activase (EC 6.3.4.-)

C;Species: Proteus mirabilis  
C;Date: 28-Aug-1985 #sequence\_revision 07-Oct-1994 #text\_change 02-Feb-2001  
C;Accession: JN0755; S32046  
R;Sriwanthana, B.; Island, M.D.; Mobley, H.L.T.  
Gene 129, 103-106, 1993

A;Title: Sequence of the Proteus mirabilis urease accessory gene ureG.  
A;Reference number: JN0755; MUID:93328109; PMID:8335248  
A;Accession: JN0755

A;Molecule type: DNA

A;Residues: 1-205 <SRI>

A;Cross-references: EMBL:Z21940; NID:g287737; PIDN:CAA79934.1; PID:g287738  
C;Comment: This protein may play a chaperonin-like role in the insertion of nickel

C;Genetics:

A;Gene: ureG

C;Superfamily: hydrogenase expression/formation protein hypB

C;Keywords: GTP binding; ligase; molecular chaperone; nucleotide binding; P-loop

F;14-21/Region: nucleotide-binding motif A (P-loop)

F;151-154/Region: GTP-binding NKXD motif

Query Match 31.2%; Score 5; DB 2; Length 205;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
|  
|  
|  
|  
Db 194 IDFII 198

RESULT 33

AC1787

thymidylate kinase homolog lln2841 [imported] - Listeria innocua (strain Clip11262)

C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C;Accession: AC1787

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; B  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsi  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam,  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; We  
A;Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AC1787

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-208 <GLA>

A;Cross-references: GB:AL592022; PIDN:CAC98067.1; PID:gl6415376; GSPDB:GN00178  
A;Experimental source: strain Clip11262

C;Genetics:

A;Gene: lln2841

C;Superfamily: dTMP kinase

Query Match 31.2%; Score 5; DB 2; Length 208;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|  
|  
|  
|  
Db 31 GIDFI 35

RESULT 34  
AD1411

thymidylate kinase homolog lmo2693 [imported] - Listeria monocytogenes (strain EGD-

C;Species: Listeria monocytogenes

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C;Accession: AD1411

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; B

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsi

D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam,

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; We

A;Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AD1411

A;Status: preliminary

A;Molecule type: DNA



A;Residues: 1-208 <GLA>  
A;Cross-references: GB:NC\_003210; PIDN:CAD00906.1; PID:g164112193; GSPDB:GN00177  
A;Experimental source: strain EGD-e  
C;Genetics:  
A;Gene: lmo2693  
C;Superfamily: dTMP kinase

Query Match 31.2%; Score 5; DB 2; Length 208;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|  
|  
|  
|  
Db 31 GIDFI 35

RESULT 35  
AC2982

RhtB family transporter rhtB [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C;Accession: AC2982  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: AB2577; PMID:11743193

A;Accession: AC2982

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-210 <KUR>

A;Cross-references: GB:AE008689; PIDN:AAL44273.1; PID:g17741859; GSPDB:GN00187

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: rhtB

A;Map position: linear chromosome

Query Match 31.2%; Score 5; DB 2; Length 210;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|  
|  
|  
|  
Db 159 GIDFI 163

RESULT 36  
T22896

hypothetical protein F58B3.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000

C;Accession: T22896

R;Harris, B.

submitted to the EMBL Data Library, May 1996

A;Reference number: Z19633

A;Accession: T22896

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-214 <WIL>

A;Cross-references: EMBL:Z73427; PIDN:CAA97801.1; GSPDB:GN00022; CESP:F58B3.3

A;Experimental source: clone F58B3

C;Genetics:

A;Gene: CESP:F58B3.3

A;Map position: 4

A;Introns: 68/1

C;Superfamily: Caenorhabditis elegans hypothetical protein F58B3.3

Query Match

Best Local Similarity 31.2%; Score 5; DB 2; Length 214;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|  
|  
|  
|  
Db 20 GIDFI 24

RESULT 37

T22892

hypothetical protein F58B3.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000

C;Accession: T22892

R;Harris, B.

submitted to the EMBL Data Library, May 1996

A;Reference number: Z19633

A;Accession: T22892

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-214 <WIL>

A;Cross-references: EMBL:Z73427; PIDN:CAA97797.1; GSPDB:GN00022; CESP:F58B3.1

A;Experimental source: clone F58B3

C;Genetics:

A;Gene: CESP:F58B3.1

A;Map position: 4

A;Introns: 68/1

C;Superfamily: Caenorhabditis elegans hypothetical protein F58B3.3

Query Match 31.2%; Score 5; DB 2; Length 214;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|  
|  
|  
|  
Db 20 GIDFI 24

RESULT 38

D81792

conserved hypothetical protein NMA2192 [imported] - Neisseria meningitidis (stra  
C;Species: Neisseria meningitidis

C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C;Accession: D81792

R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.

; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Raj

Nature 404, 502-506, 2000

A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis

A;Reference number: A81775; MUID:20222556; PMID:10761919

A;Accession: D81792

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-219 <PAR>

A;Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85404.1; PID

A;Experimental source: serogroup A, strain Z2491

C;Genetics:

A;Gene: NMA2192

Query Match 31.2%; Score 5; DB 2; Length 219;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|  
|  
|  
|  
Db 189 GIDFI 193

RESULT 39

RWHU28

T-cell surface glycoprotein CD28 precursor - human

C;Species: Homo sapiens (man)

C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 22-Jun-1999

C;Accession: A39983; A45895

R;Aruffo, A.; Seed, B.

Proc. Natl. Acad. Sci. U.S.A. 84, 8573-8577, 1987

A;Title: Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system  
A;Reference number: A39983; MUID:88068631; PMID:2825196  
A;Accession: A39983  
A;Molecule type: mRNA  
A;Residues: 1-220 <AR>  
A;Cross-references: GB:J02988; NID:g338444; PIDN:AAA60581.1; PID:g338445  
R;Lee, K.P.; Taylor, C.; Petryniak, B.; Turka, L.A.; June, C.H.; Thompson, C.B.  
J. Immunol. 145, 344-352, 1990  
A;Title: The genomic organization of the CD28 gene. Implications for the regulation of CD28 gene expression  
A;Reference number: A45895; MUID:90293482; PMID:2162892  
A;Accession: A45895  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 35-134 <LEE>  
A;Cross-references: GB:M37813  
C;Genetics:  
A;Gene: GDB:CD28  
A;Cross-references: GDB:118765; OMIM:186760  
A;Map position: 2q33-2q34  
C;Superfamily: T-cell surface glycoprotein CD28; immunoglobulin homology  
C;Keywords: glycoprotein; homodimer; T-cell; transmembrane protein  
F;1-18/Domain: signal sequence #status predicted <SIG>  
F;19-220/Product: T-cell surface glycoprotein CD28 #status predicted <MAT>  
F;19-152/Domain: extracellular #status predicted <EXT>  
F;33-114/Domain: immunoglobulin homology <IMM>  
F;153-179/Domain: transmembrane #status predicted <TMM>  
F;180-220/Domain: intracellular #status predicted <INT>  
F;71,92,105,129/Binding.site: carbohydrate (Asn) (covalent) #status predicted

Query Match 31.2%; Score 5; DB 1; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIIFW 13  
|||||  
Db 174 FIIIFW 178

RESULT 40  
B90393  
hypothetical protein SSO2231 [imported] - Sulfolobus solfataricus  
C;Species: Sulfolobus solfataricus  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C;Accession: B90393  
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.  
arrett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A;Description: Sulfolobus solfataricus complete genome.  
A;Reference number: A99139  
A;Accession: B90393  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-220 <KUR>  
A;Cross-references: GB:AE006641; NID:g13815533; PIDN:AAK42401.1; GSPDB:GN00155  
C;Genetics:  
A;Gene: SSO2231

Query Match 31.2%; Score 5; DB 2; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12  
|||||  
Db 164 DFIIIF 168

RESULT 41  
C64201  
ribosomal protein S6 modification protein rimK homolog - Mycoplasma genitalium  
C;Species: Mycoplasma genitalium  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 07-Dec-1999  
C;Accession: C64201

R;Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann  
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Meri-  
C.A.; Venter, J.C.  
Science 270, 397-403, 1995  
A;Title: The minimal gene complement of Mycoplasma genitalium.  
A;Reference number: A64200; MUID:96026346; PMID:7569993  
A;Accession: C64201  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-225 <TIGR>  
A;Cross-references: GB:U39679; GB:L43967; NID:g1045668; PID:g1045680; TIGR:MG012  
A;Experimental source: strain G-37  
C;Genetics:  
A;Genetic code: SGC3  
C;Superfamily: ribosomal protein S6 modification protein rimK

Query Match 31.2%; Score 5; DB 1; Length 225;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGIDF 9  
|||||  
Db 183 CGIDF 187

RESULT 42  
AC0587  
KDP operon transcription regulatory protein STY0743 [imported] - Salmonella enterica  
C;Species: Salmonella enterica subsp. enterica serovar Typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AC0587  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Ch-  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.;  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Steve-  
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica  
A;Reference number: AB0502; PMID:11677608  
A;Accession: AC0587  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-225 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD05165.1; PID:g16501937; GSPDB:GN00176  
C;Genetics:  
A;Gene: STY0743  
C;Superfamily: ompR protein; response regulator homology

Query Match 31.2%; Score 5; DB 2; Length 225;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|||||  
Db 60 GIDFI 64

RESULT 43  
H97331  
response regulator (Chey-like receiver domain and HTH-type DNA-binding domain) [J-  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C;Accession: H97331  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacter  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: H97331  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-230 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK81443.1; PID:g15026610; GSPDB:GN00168

Search completed: May 11, 2003, 20:13:43  
Job time : 7.95575 secs

A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC3517  
C;Superfamily: ompR protein; response regulator homology

Query Match 31.2%; Score 5; DB 2; Length 230;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11  
|  
|  
|  
|  
Db 47 IDFI 51

RESULT 44

T32999  
hypothetical protein F17E9.11 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 20-Jun-2000  
C;Accession: T32999  
R;Woessner, J.  
submitted to the EMBL Data Library, February 1998  
A;Description: The sequence of C. elegans cosmid F17E9.  
A;Reference number: 221262  
A;Accession: T32999  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-230 <WOE>  
A;Cross-references: EMBL:AF047656; PIDN:AAC05110.1; GSPDB:GN00022; CESP:F17E9.11  
A;Experimental source: strain Bristol N2; clone F17E9  
C;Genetics:  
A;Gene: CESP:F17E9.11  
A;Map position: 4  
C;Superfamily: Caenorhabditis elegans hypothetical protein F58B3.3

Query Match 31.2%; Score 5; DB 2; Length 230;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|  
|  
|  
|  
Db 20 GIDFI 24

RESULT 45

AF0367  
conserved hypothetical protein YPO3023 [imported] - Yersinia pestis (strain CO92)  
C;Species: Yersinia pestis  
C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C;Accession: AF0367  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001  
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A;Reference number: AB0001; MUID:21470413; PMID:11586360  
A;Accession: AF0367  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-233 <KUR>  
A;Cross-references: GB:AL590842; PIDN:CAC92265.1; PID:g15980976; GSPDB:GN00175  
C;Genetics:  
A;Gene: YPO3023

Query Match 31.2%; Score 5; DB 2; Length 233;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|  
|  
|  
|  
Db 144 GIDFI 148





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:01:37 ; Search time 3.25664 Seconds  
(without alignments)  
203.775 Million cell updates/sec

Title: US-09-854-133-587

Perfect score: 16

Sequence: 1 FQANGGIDFIIFWIFW 16

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	37.5	283	1 NADC_METJA	Q57916 methanococ
2	6	37.5	372	1 NU1C_PLEBO	Q00242 plectonema
3	6	37.5	590	1 NU5M_TRYBB	P04540 trypanosoma
4	6	37.5	674	1 POL_CAMVD	P03556 cauliflower
5	6	37.5	679	1 POL_CAMVC	P03555 cauliflower
6	6	37.5	679	1 POL_CAMVE	Q02964 cauliflower
7	6	37.5	679	1 POL_CAMVS	P03554 cauliflower
8	6	37.5	680	1 POL_CAMVN	Q00962 cauliflower
9	5	31.2	91	1 YC19_CYACA	Q9tm45 cyanidium c
10	5	31.2	105	1 C55L_SYNY3	P42351 synechocyst
11	5	31.2	120	1 MSCL_STAAM	O68285 staphylococ
12	5	31.2	128	1 MSCL_HABIN	P44789 haemophilus
13	5	31.2	135	1 HV02_XENLA	P20957 xenopus lae
14	5	31.2	148	1 AROQ_PASMU	P57903 pasteurella
15	5	31.2	170	1 HEMX_PROMI	Q51887 proteus mir
16	5	31.2	195	1 ENGB_LACLA	Q9cge5 lactococcus
17	5	31.2	195	1 ENGB_LACLC	Q91691 lactococcus
18	5	31.2	195	1 YPEC_LACLC	Q04731 lactococcus
19	5	31.2	205	1 UREG_PROMI	Q06206 proteus mir
20	5	31.2	220	1 CD28_HUMAN	P10747 homo sapien
21	5	31.2	235	1 SCE2_YEAST	Q03878 saccharomyc
22	5	31.2	248	1 SCOA_MYCTU	O06167 mycobacteri
23	5	31.2	264	1 AQPA_HUMAN	Q96ps8 homo sapien
24	5	31.2	268	1 TRPA_PSEAE	P07344 pseudomonas
25	5	31.2	269	1 TRPA_PSEPU	P11081 pseudomonas
26	5	31.2	269	1 TRPA_PSESY	P34816 pseudomonas
27	5	31.2	287	1 Y012_MYCGE	P47258 mycoplasma
28	5	31.2	288	1 Y012_MYCPN	P75097 mycoplasma
29	5	31.2	307	1 PPP4_HUMAN	P33172 homo sapien
30	5	31.2	307	1 PPP4_RABIT	P11084 oryctolagus
31	5	31.2	310	1 YSI9_STRCO	Q9zbi5 streptomyce
32	5	31.2	320	1 YC57_ARCFU	Q29011 archaeoglob
33	5	31.2	321	1 MRAY_ENTFA	O07107 enterococcu

34	5	31.2	330	1 MCE_CHVP1	Q84424 paramecium
35	5	31.2	337	1 AT9B_HUMAN	O43861 homo sapien
36	5	31.2	337	1 YHJD_ECOLI	P37642 escherichia
37	5	31.2	338	1 SYFA_AQUAE	O67087 aquifex ae
38	5	31.2	341	1 CYS1_OSTOS	P25802 oostertagia
39	5	31.2	342	1 CYS1_HAECO	P19092 haemonchus
40	5	31.2	342	1 CYS2_HAECO	P25793 haemonchus
41	5	31.2	342	1 HEM2_RICPR	Q9zc84 rickettsia
42	5	31.2	342	1 Y762_METJA	Q58172 methanococc
43	5	31.2	343	1 AROB_HELPJ	Q9zmf2 helicobacte
44	5	31.2	343	1 AROB_HELPY	P56081 helicobacte
45	5	31.2	351	1 AROB_CAMJE	Q9pnt2 campylobact
46	5	31.2	354	1 US28_HCMVA	P09704 human cytom
47	5	31.2	356	1 AROB_THETN	Q8rb14 thermoanaer
48	5	31.2	362	1 AROB_BACSU	P31102 bacillus su
49	5	31.2	380	1 FTZ2_METJA	Q58039 methanococc
50	5	31.2	388	1 ARRC_HUMAN	P36575 homo sapien
51	5	31.2	400	1 NIFV_ENTAG	Q52070 enterobacte
52	5	31.2	401	1 T11B_RAT	O08727 rattus norv
53	5	31.2	404	1 EX7L_FUSNN	Q8ren3 fusobacteri
54	5	31.2	407	1 DEOB_ECOLI	P07651 escherichia
55	5	31.2	412	1 SYV_CHLPN	Q9z8i2 chlamydia p
56	5	31.2	416	1 CINA_BACSU	P46323 bacillus su
57	5	31.2	419	1 O56A_DROME	Q9v8y7 drosophila
58	5	31.2	439	1 PPNK_MOUSE	P58058 mus musculu
59	5	31.2	446	1 PPNK_HUMAN	Q95544 homo sapien
60	5	31.2	469	1 DNAA_TREDE	O87546 treponema d
61	5	31.2	483	1 ACPA_BACAN	Q44643 bacillus an
62	5	31.2	515	1 CALR_MOUSE	Q60755 mus musculu
63	5	31.2	517	1 FU26_YEAST	P31381 saccharomyc
64	5	31.2	549	1 COX1_TRYBB	P04371 trypanosoma
65	5	31.2	587	1 KPYK_BACST	Q02499 bacillus st

# ALIGNMENTS

## RESULT 1

ID	NADC_METJA	STANDARD;	PRT;	283 AA.
AC	Q57916;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Probable nicotinate-nucleotide pyrophosphorylase [carboxylating]			
DE	(EC 2.4.2.19) (Quinolinatase phosphoribosyltransferase			
DE	[decarboxylating]) (QAPRTase).			
GN	NADC OR MJ0493.			
OS	Methanococcus jannaschii.			
OC	Archaea; Euryarchaeota; Methanococci; Methanococcales;			
OC	Methanocaldococcaceae; Methanocaldococcus.			
OX	NCBI_TaxID=2190;			
RP	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=JAL-1 / DSM 2661 / ATCC 43067;			
RX	MEDLINE=96337999; PubMed=8688087;			
RA	Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,			
RA	Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,			
RA	Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,			
RA	Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,			
RA	Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,			
RA	Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,			
RA	Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,			
RA	Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;			
RT	"Complete genome sequence of the methanogenic archaeon, Methanococcus			
RT	jannaschii."			
RL	Science 273:1058-1073(1996).			
CC	-!- CATALYTIC ACTIVITY: Nicotinate D-ribonucleotide + diphosphate +			
CC	CO(2) = pyridine-2,3-dicarboxylate + 5-phospho-alpha-D-ribose 1-			
CC	diphosphate.			
CC	-!- PATHWAY: NAD biosynthesis; aspartate to NAMN; third (last) step.			
CC	-!- SIMILARITY: BELONGS TO THE NADC/MODD FAMILY.			

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U67499; AAB98483.1; -  
DR HSSP; O06594; 1QPO.  
DR TIGR; MJ0493; -  
DR InterPro; IPR004393; NadC.  
DR InterPro; IPR002638; QRPTase.  
DR Pfam; PF01729; QRPTase; 1.  
DR Pfam; PF02749; QRPTase\_N; 1.  
DR ProDom; PD003988; QRPTase; 1.  
DR TIGRFAMs; TIGR00078; nadC; 1.  
KW pyridine nucleotide biosynthesis; Transferase; Glycosyltransferase;  
KW Complete proteome.  
SQ SEQUENCE 283 AA; 31970 MW; BDD118E9CE1401A8 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 283;  
Best Local Similarity 100.0%; Pred. No. 5.8;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGIDFI 10  
Db 48 CGIDFI 53

RESULT 2  
NUIC\_PLEBO STANDARD; PRT; 372 AA.  
ID NUIC\_PLEBO  
AC Q00242;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE NADH-plastoquinone oxidoreductase chain 1 homolog (EC 1.6.5.3).  
GN NDH1.  
OS Plectonema boryanum.  
OC Bacteria; Cyanobacteria; Oscillatoriales; Plectonema.  
OX NCBI\_TaxID=1184;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-IAM M-101;  
RA Takahashi Y., Shonai F., Fujita Y., Kohchi T., Ohyaama K.,  
RA Matsubara H.;  
RT "Structure of a co-transcribed gene cluster, ndh1-frxB-ndh6-ndh4L,  
RT cloned from the filamentous cyanobacterium Plectonema boryanum."  
RL Plant Cell Physiol. 32:969-981(1991).  
CC -!- CATALYTIC ACTIVITY: NADH + plastoquinone = NAD(+) + plastoquinol.  
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; D01014; BAA00814.1; -  
DR PIR; JQ2135; JQ2135.  
DR InterPro; IPR001694; Resp\_NADH\_dh1.  
DR Pfam; PF00146; NADHdh; 1.  
DR PROSITE; PS00667; COMPLEX1\_ND1\_1; FALSE\_NEG.  
DR PROSITE; PS00668; COMPLEX1\_ND1\_2; 1.  
KW Oxidoreductase; NAD; Plastoquinone; Transmembrane.  
SQ SEQUENCE 372 AA; 40239 MW; 971729ED98F2094D CRC64;

Query Match 37.5%; Score 6; DB 1; Length 372;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFWI 14  
Db 216 FIIFWI 221

RESULT 3

NU5M\_TRYBB STANDARD; PRT; 590 AA.  
ID NU5M\_TRYBB  
AC P04540;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3).  
GN ND5.  
OS Trypanosoma brucei brucei.  
OG Mitochondrion.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85037915; PubMed=6093040;  
RA Hensgens L.A.M., Brakenhoff J., de Vries B.F., Sloof P., Tromp M.C.,  
RA van Boom J.H., Benne R.;  
RT "The sequence of the gene for cytochrome c oxidase subunit I, a  
RT frameshift containing gene for cytochrome c oxidase subunit II and  
RT seven unassigned reading frames in Trypanosoma brucei mitochondrion  
RT maxi-circle DNA."  
RL Nucleic Acids Res. 12:7327-7344(1984).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=164;  
RX MEDLINE=87201680; PubMed=2437452;  
RA Jasmer D.P., Feagin J.E., Payne M., Stuart K.;  
RT "Variation of G-rich mitochondrial transcripts among stocks of  
RT Trypanosoma brucei."  
RL Mol. Biochem. Parasitol. 22:259-272(1987).  
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X01094; CAB57807.1; -  
DR EMBL; M14820; AAB59225.1; -  
DR PIR; A04519; QQUTC5.  
DR InterPro; IPR001750; Oxidored\_q1.  
DR Pfam; PF00361; oxidored\_q1; 1.  
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.  
SQ SEQUENCE 590 AA; 71494 MW; CAE231AD9DFF5282 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 590;  
Best Local Similarity 100.0%; Pred. No. 9.9;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWIF 15  
Db 115 IIFWIF 120

RESULT 4

POL\_CAMVD STANDARD; PRT; 674 AA.  
ID POL\_CAMVD  
AC P03556;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Enzymatic polyprotein [Contains: Aspartic protease (EC 3.4.23.-);  
DE Endonuclease; Reverse transcriptase (EC 2.7.7.49)].

GN V. Cauliflower mosaic virus (strain D/H) (CaMV).  
 OS Viruses; Retrovirus; Caulimoviridae; Caulimovirus.  
 OC NCBI\_TaxID=10645;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=83106468; PubMed=7152260;  
 RX Balazs E., Guilley H., Jonard G., Richards K.;  
 RA "Nucleotide sequence of DNA from an altered-virulence isolate D/H of  
 RT the cauliflower mosaic virus.";  
 RL Gene 19:239-249(1982).  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
 CC + {DNA}(N).  
 CC -1- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.  
 CC -1- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M10376; AAA46350.1; ALT\_INIT.  
 CC PIR; A04159; QQCV5.  
 CC MEROPS; A03.001; -.  
 CC InterPro; IPR000588; Peptidase\_A3.  
 CC InterPro; IPR000477; RVTse.  
 CC Pfam; PF00078; rvt; 1.  
 CC Pfam; PF02160; Peptidase\_A3; 1.  
 CC PRINTS; PR00731; CAULIMOPTASE.  
 CC Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;  
 CC Transferase; Endonuclease; Polyprotein.  
 CC ACT\_SITE 47 47  
 CC SIMILAR 40 130 TO RETROVIRUS GAG/POL PROTEASE DOMAIN.  
 CC SIMILAR 260 620 TO RETROVIRUS GAG/POL DNA POLYMERASE  
 CC DOMAIN.  
 CC SEQUENCE 674 AA; 78164 MW; E04E3222D349E29 CRC64;  
 DR PIR; A04159; QQCV5.  
 DR MEROPS; A03.001; -.  
 DR InterPro; IPR000588; Peptidase\_A3.  
 DR InterPro; IPR000477; RVTse.  
 DR Pfam; PF00078; rvt; 1.  
 DR Pfam; PF02160; Peptidase\_A3; 1.  
 DR PRINTS; PR00731; CAULIMOPTASE.  
 DR Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;  
 DR Transferase; Endonuclease; Polyprotein.  
 DR ACT\_SITE 47 47  
 DR SIMILAR 40 130 TO RETROVIRUS GAG/POL PROTEASE DOMAIN.  
 DR SIMILAR 260 620 TO RETROVIRUS GAG/POL DNA POLYMERASE  
 DR DOMAIN.  
 DR SEQUENCE 674 AA; 78164 MW; E04E3222D349E29 CRC64;  
 Query Match 37.5%; Score 6; DB 1; Length 674;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 GIDFII 11  
 DB 111 GIDFII 116  
 RESULT 5  
 POL\_CAMVC STANDARD; PRT; 679 AA.  
 ID POL\_CAMVC  
 AC P03555;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Enzymatic polyprotein [Contains: Aspartic protease (EC 3.4.23.-);  
 DE Endonuclease; Reverse transcriptase (EC 2.7.7.49)].  
 GN V.  
 OS Cauliflower mosaic virus (strain CM-1841) (CaMV).  
 OC Viruses; Retrovirus; Caulimoviridae; Caulimovirus.  
 OX NCBI\_TaxID=10644;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=82014878; PubMed=6269062;  
 RA Gardner R.C., Howarth A.J., Hahn P., Brown-Luedi M., Shepherd R.J.,  
 RA Messing J.;  
 RT "The complete nucleotide sequence of an infectious clone of  
 RT cauliflower mosaic virus by M13mp7 shotgun sequencing.";  
 RL Nucleic Acids Res. 9:2871-2888(1981).  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
 CC + {DNA}(N).

CC -1- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.  
 CC -1- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; V00140; -; NOT\_ANNOTATED\_CDS.  
 CC PIR; A04159; QQCV5.  
 CC MEROPS; A03.001; -.  
 CC InterPro; IPR000588; Peptidase\_A3.  
 CC InterPro; IPR000477; RVTse.  
 CC Pfam; PF00078; rvt; 1.  
 CC Pfam; PF02160; Peptidase\_A3; 1.  
 CC PRINTS; PR00731; CAULIMOPTASE.  
 CC Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;  
 CC Transferase; Endonuclease; Polyprotein.  
 CC ACT\_SITE 45 45  
 CC SIMILAR 40 130 TO RETROVIRUS GAG/POL PROTEASE DOMAIN.  
 CC SIMILAR 260 620 TO RETROVIRUS GAG/POL DNA POLYMERASE  
 CC DOMAIN.  
 CC SEQUENCE 679 AA; 78669 MW; E21F8BC528C9DB8D CRC64;  
 Query Match 37.5%; Score 6; DB 1; Length 679;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 GIDFII 11  
 DB 109 GIDFII 114  
 RESULT 6  
 POL\_CAMVC STANDARD; PRT; 679 AA.  
 ID POL\_CAMVC  
 AC Q02964;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Enzymatic polyprotein [Contains: Aspartic protease (EC 3.4.23.-);  
 DE Endonuclease; Reverse transcriptase (EC 2.7.7.49)].  
 GN V.  
 OS Cauliflower mosaic virus (strain BBC) (CaMV).  
 OC Viruses; Retrovirus; Caulimoviridae; Caulimovirus.  
 OX NCBI\_TaxID=31556;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93154593; PubMed=8428667;  
 RA Chenault K.D., Melcher U.K.;  
 RA "The complete nucleotide sequence of cauliflower mosaic virus isolate  
 RT BBC.";  
 RL Gene 123:255-257(1993).  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
 CC + {DNA}(N).  
 CC -1- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.  
 CC -1- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M90542; AAA62375.1; -.  
 CC MEROPS; A03.001; -.

```
DR InterPro; IPR000588; Peptidase_A3.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF02160; Peptidase_A3; 1.
DR PRINTS; PR00731; CAULIMOPTASE.
KW Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;
KW Transferase; Endonuclease; Polyprotein.
FT ACT_SITE 45 45
FT SIMILAR 40 130
FT SIMILAR 260 620
FT TO RETROVIRUS GAG/POL PROTEASE DOMAIN.
FT TO RETROVIRUS GAG/POL DNA POLYMERASE
SQ SEQUENCE 679 AA; 78713 MW; 2D8031B62F1CB05A CRC64;

Query Match
Best Local Similarity 37.5%; Score 6; DB 1; Length 679;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11
Db 109 GIDFII 114

RESULT 7
POL_CAMVS
ID POL_CAMVS STANDARD; PRT; 679 AA.
AC P03554;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Enzymatic polyprotein [Contains: Aspartic protease]
DE Endonuclease; Reverse transcriptase (EC 2.7.7.49)].
GN V.
OS Cauliflower mosaic virus (strain Strasbourg) (CamV).
OC Viruses; Retroviral viruses; Caulimoviridae; Caulimovirus.
OX NCBI_TaxID=10648;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81001865; PubMed=7407912;
RA Franck A., Guille H., Jonard G., Richards K., Hirth L.;
RT "Nucleotide sequence of cauliflower mosaic virus DNA.";
RL Cell 21:285-294(1980).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}(N).
CC -!- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.
CC -!- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V00141; CAA23460.1; -.
CC PIR; A04159; QOCV5.
CC MEROPS; A03.001; -.
CC InterPro; IPR000588; Peptidase_A3.
CC InterPro; IPR000477; RVTse.
CC Pfam; PF00078; rvt; 1.
CC Pfam; PF02160; Peptidase_A3; 1.
CC PRINTS; PR00731; CAULIMOPTASE.
CC Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;
CC Transferase; Endonuclease; Polyprotein.
CC ACT_SITE 45 45
CC SIMILAR 40 130
CC SIMILAR 260 620
CC TO RETROVIRUS GAG/POL PROTEASE DOMAIN.
CC TO RETROVIRUS GAG/POL DNA POLYMERASE
SQ SEQUENCE 679 AA; 78629 MW; 9EE527BCD460B766 CRC64;

Query Match
Best Local Similarity 37.5%; Score 6; DB 1; Length 679;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11
Db 109 GIDFII 114
```

```
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11
Db 109 GIDFII 114

RESULT 8
POL_CAMVN
ID POL_CAMVN STANDARD; PRT; 680 AA.
AC Q00962;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Enzymatic polyprotein [Contains: Aspartic protease (EC 3.4.23.-);
DE Endonuclease; Reverse transcriptase (EC 2.7.7.49)].
GN V.
OS Cauliflower mosaic virus (strain NY8153) (CamV).
OC Viruses; Retroviral viruses; Caulimoviridae; Caulimovirus.
OX NCBI_TaxID=31557;
RN [1]
RP SEQUENCE FROM N.A.
RA Chenault K.D., Steffens D.L., Melcher U.K.;
RT "Nucleotide sequence of cauliflower mosaic virus isolate NY8153.";
RL Plant Physiol. 100:542-545(1992).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}(N).
CC -!- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.
CC -!- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M90541; AAA46358.1; -.
CC MEROPS; A03.001; -.
CC InterPro; IPR000588; Peptidase_A3.
CC InterPro; IPR000477; RVTse.
CC Pfam; PF00078; rvt; 1.
CC Pfam; PF02160; Peptidase_A3; 1.
CC PRINTS; PR00731; CAULIMOPTASE.
CC Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;
CC Transferase; Endonuclease; Polyprotein.
CC ACT_SITE 46 46
CC SIMILAR 41 131
CC SIMILAR 261 621
CC TO RETROVIRUS GAG/POL PROTEASE DOMAIN.
CC TO RETROVIRUS GAG/POL DNA POLYMERASE
SQ SEQUENCE 680 AA; 78665 MW; FCE02E09647C2221 CRC64;

Query Match
Best Local Similarity 37.5%; Score 6; DB 1; Length 680;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11
Db 110 GIDFII 115

RESULT 9
YC19_CYACA
ID YC19_CYACA STANDARD; PRT; 91 AA.
AC Q9TM45;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 10.5 kDa protein ycf19.
GN YCF19 OR YCF49.
OS Cyanidium caldarium.
```



entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

OG Chloroplast.  
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;  
OC Cyanidium.  
OX NCBI\_TaxID=2771;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RK-1;  
RX MEDLINE=20496959; PubMed=11040290;  
RA Gloeckner G., Rosenthal A., Valentin K.-U.;  
RT "The structure and gene repertoire of an ancient red algal plastid genome";  
RL J. Mol. Evol. 51:382-390(2000).  
CC -1- SIMILARITY: BELONGS TO THE YCF19 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF022186; AAF13024.1; -.  
DR InterPro; IPR003425; Unk\_YGGT.  
DR Pfam; PF02325; YGGT; 1.  
KW Chloroplast; Hypothetical protein.  
SQ SEQUENCE 91 AA; 10510 MW; B998BDA83208D00E CRC64;

Query Match 31.2%; Score 5; DB 1; Length 91;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11  
Db 77 IDFI 81

RESULT 10  
C55L\_SYNY3  
ID C55L\_SYNY3 STANDARD; PRT; 105 AA.

AC P42351;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Cytochrome c-553-like precursor.  
GN CYTM OR SLL1245.  
OS Synecocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.  
OX NCBI\_TaxID=1148;  
RN [1]

RA Malakhov M.P., Wada H., Los D.A., Semenenko V.E., Murata N.;  
RT "A new type of cytochrome c from Synecocystis PCC6803";  
RL J. Plant Physiol. 144:259-264(1994).  
RN [2]

RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,  
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,  
RA Yamada M., Yasuda M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions";  
RL DNA Res. 3:109-136(1996).

-1- PTM: BINDS ONE HEME GROUP PER MOLECULE.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC EMBL; D10716; BAA01559.1; -.  
CC EMBL; D90912; BAA18172.1; ALT\_INIT.  
CC HSP; P00080; LHRO.  
DR InterPro; IPR000345; CytC\_heme\_bind.  
DR InterPro; IPR003088; Cyt\_C1.  
DR Pfam; PF00034; cytochrome\_c; 1.  
DR ProDom; PD004020; Cyt\_C1C; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
KW Electron transport; Photosynthesis; Heme; Signal; Complete proteome.  
FT SIGNAL 1 29 POTENTIAL.  
FT CHAIN 30 105 CYTOCHROME C-553-LIKE.  
FT BINDING 45 45 HEME (COVALENT) (BY SIMILARITY).  
FT BINDING 48 48 HEME (COVALENT) (BY SIMILARITY).  
FT METAL 49 49 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
FT METAL 85 85 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
SQ SEQUENCE 105 AA; 11458 MW; 21247DEF9D008A94 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 105;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FQANC 5  
Db 41 FQANC 45

RESULT 11  
MSCL\_STAAM  
ID MSCL\_STAAM STANDARD; PRT; 120 AA.  
AC O68285;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Large-conductance mechanosensitive channel.  
GN MSCL OR SAV1347 OR SAIL182 OR MW1235.  
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),  
OS Staphylococcus aureus (strain N315),  
OS Staphylococcus aureus (strain MW2), and  
OS Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=158878, 158879, 196620, 1280;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=Mu50 / ATCC 700699, and N315;  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
RT aureus";  
RL Lancet 357:1225-1240(2001).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=RW2;  
RX MEDLINE=22040717; PubMed=12044378;  
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,  
RA Yamamoto K., Hiramatsu K.;  
RT "Genome and virulence determinants of high virulence community-  
RT acquired MRSA";  
RL Lancet 359:1819-1827(2002).  
RN [3]

RP SEQUENCE FROM N.A.  
RC STRAIN=RN450;  
RX MEDLINE=98294049; PubMed=9632260;  
Moe P.C., Blount P., Kung C.;

RT Functional and structural conservation in the mechanosensitive  
channel mscL implicates elements crucial for mechanosensation."  
RL Mol. Microbiol. 28:583-592(1998).  
CC -|- FUNCTION: CHANNEL THAT OPENS IN RESPONSE TO STRETCH FORCES IN THE  
MEMBRANE LIPID BILAYER. MAY PARTICIPATE IN THE REGULATION OF  
CC OSMOTIC PRESSURE CHANGES WITHIN THE CELL.  
CC -|- SUBUNIT: HOMOPENTAMER (BY SIMILARITY).  
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -|- SIMILARITY: BELONGS TO THE MSC\_L FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AP003362; BAB57509.1; -  
DR EMBL; AP003133; BAB42440.1; -  
DR EMBL; AP004826; BAB95100.1; -  
DR EMBL; AF029731; AAC38560.1; -  
DR HSSP; O53898; IMSL.  
DR InterPro; IPR001185; MS\_channel.  
DR Pfam; PF01741; MscL; 1.  
DR PRINTS; PR01264; MECHCHANNEL.  
DR ProDom; PD007253; MS\_channel; 1.  
DR TIGRFAMS; TIGR00220; mscL; 1.  
DR PROSITE; PS01327; MscL; 1.  
KW Transmembrane; Ionic channel; Complete proteome.  
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 14 40 BY SIMILARITY.  
FT DOMAIN 41 62 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 63 84 BY SIMILARITY.  
FT DOMAIN 85 120 CYTOPLASMIC (POTENTIAL).  
SQ SEQUENCE 120 AA; 13616 MW; A4D1E6B2A7B7D2E5 CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 120;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 7 IDFII 11  
Db 71 IDFII 75

RESULT 12  
MSCL\_HAEIN STANDARD; PRT; 128 AA.  
ID MSCL\_HAEIN STANDARD; PRT; 128 AA.  
AC P44789;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Large-conductance mechanosensitive channel.  
GN MscL OR HI0626.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae

Rd.";  
RL Science 269:496-512(1995).  
CC -|- FUNCTION: CHANNEL THAT OPENS IN RESPONSE TO STRETCH FORCES IN THE  
MEMBRANE LIPID BILAYER. MAY PARTICIPATE IN THE REGULATION OF  
CC OSMOTIC PRESSURE CHANGES WITHIN THE CELL (BY SIMILARITY).  
CC -|- SUBUNIT: HOMOPENTAMER (BY SIMILARITY).  
CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
CC (By similarity).  
CC -|- SIMILARITY: BELONGS TO THE MSC\_L FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; U32745; AAC22286.1; -  
DR HSSP; O53898; IMSL.  
DR TIGR; HI0626; -  
DR InterPro; IPR001185; MS\_channel.  
DR Pfam; PF01741; MscL; 1.  
DR PRINTS; PR01264; MECHCHANNEL.  
DR ProDom; PD007253; MS\_channel; 1.  
DR TIGRFAMS; TIGR00220; mscL; 1.  
DR PROSITE; PS01327; MscL; 1.  
KW Transmembrane; Inner membrane; Ionic channel; Complete proteome.  
FT DOMAIN 1 15 CYTOPLASMIC (BY SIMILARITY).  
FT TRANSMEM 16 42 BY SIMILARITY.  
FT DOMAIN 43 74 PERIPLASMIC (BY SIMILARITY).  
FT TRANSMEM 75 96 BY SIMILARITY.  
FT DOMAIN 97 128 CYTOPLASMIC (BY SIMILARITY).  
SQ SEQUENCE 128 AA; 14198 MW; BD3A68D27200E51A CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 128;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 7 IDFII 11  
Db 83 IDFII 87

RESULT 13  
HV02\_XENLA STANDARD; PRT; 135 AA.  
ID HV02\_XENLA STANDARD; PRT; 135 AA.  
AC P20957;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ig heavy chain V region Xig14 precursor (Fragment).  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88176921; PubMed=2451244;  
RA Schwager J., Mikoryak C.A., Steiner L.A.;  
RT "Amino acid sequence of heavy chain from Xenopus laevis IgM deduced  
from cDNA sequence: implications for evolution of immunoglobulin  
domains."  
RL Proc. Natl. Acad. Sci. U.S.A. 85:2245-2249(1988).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; J03632; AAA49791.1; -.  
DR PIR; B31933; B31933.  
DR HSSP; P01810; 2FBJ.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig; 1.  
DR SMART; SM00406; IGV; 1.  
KW Immunoglobulin V region; Signal.  
FT NON\_TER 1 1  
FT SIGNAL <1 18  
FT CHAIN 19 135 IG HEAVY CHAIN V REGION XIG14.  
FT NON\_TER 135 135  
FT SEQUENCE 135 AA; 15080 MW; EBC467105C00732E CRC64;  
SQ  
  
Query Match 31.2%; Score 5; DB 1; Length 135;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 8 DFIF 12  
Db 1 DFIF 5  
  
RESULT 14  
AROQ\_PASMU STANDARD; PRT; 148 AA.  
ID AROQ\_PASMU  
AC P57903;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 3-dehydroquinase dehydratase (EC 4.2.1.10) (3-dehydroquinase) (Type II  
DHQase).  
GN AROQ OR AROD OR PM1093.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Pasteurella.  
OC NCBI\_TaxID=747;  
CC [1]  
CC SEQUENCE FROM N.A.  
CC STRAIN=Pm70;  
CC MEDLINE=21145866; PubMed=11248100;  
CC May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
CC "Complete genomic sequence of Pasteurella multocida Pm70.";  
CC Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
CC -!- FUNCTION: Catalyzes a trans-dehydration via an enolate  
CC intermediate (By similarity).  
CC -!- CATALYTIC ACTIVITY: 3-dehydroquinase = 3-dehydroshikimate + H(2)O.  
CC -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;  
CC third step.  
CC -!- SIMILARITY: BELONGS TO THE TYPE-II 3-DEHYDROQUINASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE006150; AAK03177.1; -.  
CC HSSP; P36918; 2DHQ.  
CC InterPro; IPR001874; DHQuinase\_II.  
CC Pfam; PF01220; DHQuinase\_II; 1.  
CC TIGRFAMs; TIGR01088; aroQ; 1.  
CC PROSITE; PS01029; DEHYDROQUINASE\_II; 1.  
KW Aromatic amino acid biosynthesis; Lyase; Complete proteome.  
SQ SEQUENCE 148 AA; 16626 MW; DD4FB1CB7FB1934B CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 148;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
Db 71 IDFII 75  
  
RESULT 15  
HEMX\_PROMI STANDARD; PRT; 170 AA.  
ID HEMX\_PROMI  
AC Q51887;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Putative uroporphyrin-III C-methyltransferase (EC 2.1.1.107) (Urogen  
III methylase) (Fragment).  
GN HEMX.  
OS Proteus mirabilis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Proteus.  
OC NCBI\_TaxID=584;  
CC [1]  
CC SEQUENCE FROM N.A.  
CC MEDLINE=97028791; PubMed=8874804;  
CC Trotot P., Sismeiro O., Vivares C., Glaser P., Bresson-Roy A.,  
CC Danchin A.;  
CC "Comparative analysis of the cya locus in enterobacteria and related  
CC Gram-negative facultative anaerobes.";  
CC Biochimie 78:277-287(1996).  
CC -!- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + uroporphyrin III -  
CC 2 S-adenosyl-L-homocysteine + sirohdrochlorin.  
CC -!- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF SIROHEME AND COBALAMIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U22969; AAC44327.1; -.  
CC Porphyrin biosynthesis; Transferase; Methyltransferase.  
CC NON\_TER 170 170  
CC SEQUENCE 170 AA; 18243 MW; DCA77C4EF3E2BE2C CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 170;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 7 IDFII 11  
Db 120 IDFII 124  
  
RESULT 16  
ENGB\_LACLA STANDARD; PRT; 195 AA.  
ID ENGB\_LACLA  
AC Q9CGE5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable GTP-binding protein engB.  
GN ENGB OR LL1152.  
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OC NCBI\_TaxID=1360;  
CC [1]  
CC SEQUENCE FROM N.A.  
CC STRAIN=LL1403;  
CC MEDLINE=21235186; PubMed=11337471;  
CC Bolotin A., Wincker P., Manger S., Jaillon O., Malarre K.,  
CC Weissenbach J., Ehrlich S.D., Sorokin A.;  
CC "The complete genome sequence of the lactic acid bacterium Lactococcus  
CC lactis ssp. lactis ILL1403.";  
CC

Genome Res. 11:731-753(2001).  
-1- FUNCTION: NECESSARY FOR NORMAL CELL DIVISION AND FOR THE  
MAINTENANCE OF NORMAL SEPTATION (BY SIMILARITY).  
-1- SIMILARITY: BELONGS TO THE ENGB FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AE006347; AAK05250.1; -  
InterPro; IPR005289; GTP-binding\_dom.  
InterPro; IPR001806; Ras\_trnsmg.  
PRINTS; PR00449; RASTRNSFRMNG.  
TIGRFAMS; TIGR00650; MG442; 1.  
Cell division; Septation; GTP-binding; Complete proteome.  
NP\_BIND 32 39 GTP (POTENTIAL).  
NP\_BIND 77 81 GTP (POTENTIAL).  
NP\_BIND 144 147 GTP (POTENTIAL).  
SEQUENCE 195 AA; 22551 MW; F2750DB86AF93BFB CRC64;  
-----  
Query Match 31.2%; Score 5; DB 1; Length 195;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
QY 8 DFIIF 12  
Db 172 DFIIF 176  
-----  
RESULT 17  
ENGB\_LACLC STANDARD; PRT; 195 AA.  
ID ENGB\_LACLC  
AC Q9L6G1;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable GTP-binding protein engb.  
GN ENGB.  
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MG1363;  
RA Skinner M.M., Trempey J.E.;  
RT "Sequence analysis and comparison of the clpX region from Lactococcus  
lactis.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: NECESSARY FOR NORMAL CELL DIVISION AND FOR THE  
MAINTENANCE OF NORMAL SEPTATION (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE ENGB FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF236863; AAF63739.1; -  
InterPro; IPR005289; GTP-binding\_dom.  
InterPro; IPR001806; Ras\_trnsmg.  
PRINTS; PR00449; RASTRNSFRMNG.  
TIGRFAMS; TIGR00650; MG442; 1.  
Cell division; Septation; GTP-binding.  
NP\_BIND 32 39 GTP (POTENTIAL).  
NP\_BIND 77 81 GTP (POTENTIAL).  
NP\_BIND 144 147 GTP (POTENTIAL).  
-----

SQ SEQUENCE 195 AA; 22460 MW; F2751A8499F9250F CRC64;  
-----  
Query Match 31.2%; Score 5; DB 1; Length 195;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
QY 8 DFIIF 12  
Db 172 DFIIF 176  
-----  
RESULT 18  
YPEC\_LACLC STANDARD; PRT; 195 AA.  
ID YPEC\_LACLC  
AC Q04731;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein in PEPC 5'region (ORF 2) (Fragment).  
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AM2;  
RA MEDLINE=93175873; PubMed=8439160;  
RT Chapot-Chartier M.P., Nardi M., Chopin M.-C., Chopin A., Gripon J.-C.;  
RT "Cloning and sequencing of pepC, a cysteine aminopeptidase gene from  
Lactococcus lactis subsp. cremoris AM2.";  
RL Appl. Environ. Microbiol. 59:330-333(1993).  
CC -----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M86245; AAA74513.1; -  
PIR; A48957; A48957.  
InterPro; IPR002819; HD.  
PFam; PF01966; HD; 1.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 195 AA; 22186 MW; 2A4D27E780ABA507 CRC64;  
-----  
Query Match 31.2%; Score 5; DB 1; Length 195;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
QY 7 IDFI 11  
Db 168 IDFI 172  
-----  
RESULT 19  
UREG\_PROMI STANDARD; PRT; 205 AA.  
ID UREG\_PROMI  
AC Q06206;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Urease accessory protein ureg.  
GN UREG.  
OS Proteus mirabilis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Proteus.  
OX NCBI\_TaxID=584;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HI4320;  
RX MEDLINE=93328109; PubMed=8335248;



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

RA Sriwanthana B., Island M.D., Mobley H.L.T.;  
RT "Sequence of the Proteus mirabilis urease accessory gene ureG.";  
RL Gene 129:103-106(1993).  
CC -!- FUNCTION: PROBABLY FACILITATES NICKEL INCORPORATION.  
CC -!- SIMILARITY: BELONGS TO THE UREG FAMILY.  
CC -----

DR EMBL; J02988; AAA60581.1; -  
DR EMBL; M37815; AAA51944.1; -  
DR EMBL; M37812; AAA51944.1; JOINED.  
DR EMBL; M37813; AAA51944.1; JOINED.  
DR EMBL; M37814; AAA51944.1; JOINED.  
DR PIR; A39983; RWU28.  
DR HSSP; P16410; IAH1.  
DR Genew; HGNC:1653; CD28.  
DR MIM; 186760; -  
DR InterPro; IPR003600; Ig\_like.  
DR SMART; SM00410; IG\_like; 1.  
KW Immunoglobulin domain; T-cell; Glycoprotein; Signal; Transmembrane.  
FT SIGNAL 1 18  
FT CHAIN 19 220  
FT DOMAIN 19 152  
FT TRANSMEM 153 179  
FT DOMAIN 180 220  
FT DOMAIN 28 137  
FT CARBOHYD 37 37  
FT CARBOHYD 71 71  
FT CARBOHYD 92 92  
FT CARBOHYD 105 105  
FT CARBOHYD 129 129  
FT SEQUENCE 220 AA; 25066 MW; 1D9B6552A5878D0F CRC64;  
SQ

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; Z21940; CAA79934.1; -  
DR PIR; S32046; S32046.  
DR PIR; JN0755; JN0755.  
DR InterPro; IPR002894; HypB\_Ureg.  
DR InterPro; IPR004400; Ureg.  
DR Pfam; PF01495; HypB\_Ureg; 1.  
DR TIGRFAMs; TIGR00101; ureG; 1.  
KW Nickel; ATP-binding.  
FT NP\_BIND 14 21  
FT SEQUENCE 205 AA; 22375 MW; 9D7B7871C999636A CRC64;  
SQ

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

Query Match 31.2%; Score 5; DB 1; Length 205;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 7 IDFII 11  
Db 194 IDFII 198

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

RESULT 20  
CD28\_HUMAN STANDARD; PRT; 220 AA.  
ID CD28\_HUMAN  
AC P10747;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE T-cell-specific surface glycoprotein CD28 precursor (TP44).  
GN CD28.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88068631; PubMed=2825196;  
RA Aruffo A., Seed B.;  
RT "Molecular cloning of a CD28 cDNA by a high-efficiency COS cell  
RT expression system.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:8573-8577(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90293482; PubMed=2162892;  
RA Lee K.P., Taylor C., Petryniak B., Turka L.A., June C.H.,  
RA Thompson C.B.;  
RT "The genomic organization of the CD28 gene. Implications for the  
RT regulation of CD28 mRNA expression and heterogeneity.";  
RL J. Immunol. 145:344-352(1990).  
CC -!- FUNCTION: POSSIBLY INVOLVED IN T-CELL ACTIVATION. BINDS TO B7-1  
CC AND B7-2 (B70).  
CC -!- SUBUNIT: Homodimer; disulfide-linked. Interacts with DUSP14.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD28 entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd28.htm".  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

Query Match 31.2%; Score 5; DB 1; Length 220;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 FIFW 13  
Db 174 FIFW 178

RESULT 21  
SCE2\_YEAST STANDARD; PRT; 235 AA.  
ID SCE2\_YEAST  
AC P03878;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Intron-encoded endonuclease I-SceII (EC 3.1.-.-) (DNA endonuclease I-  
DE SceII) (DNA endonuclease AI4) (PAL4 nuclease).  
GN AI4 OR ENS2 OR I-SCEII.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OG Mitochondrion.  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D273-10B;  
RX MEDLINE=81069885; PubMed=6254986;  
RA Bonitz S.G., Coruzzi G., Thalenfeld B.E., Tzagoloff A., Macino G.;  
RT "Assembly of the mitochondrial membrane system. Structure and  
RT nucleotide sequence of the gene coding for subunit 1 of yeast  
RT cytochrome oxidase.";  
RL J. Biol. Chem. 255:11927-11941(1980).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=91016898; PubMed=2216759;  
RA Sargueil B., Hataat D., Delahodde A., Jacq C.;  
RT "In vivo and in vitro analyses of an intron-encoded DNA endonuclease  
RT from yeast mitochondria. Recognition site by site-directed  
RT mutagenesis.";  
RL Nucleic Acids Res. 18:5659-5665(1990).  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

```
CC CC -!- FUNCTION: ENDONUCLEASE INVOLVED IN INTRON HOMING. INTRODUCES A
CC CC SPECIFIC DOUBLE-STRAND BREAK AT THE JUNCTION OF THE TWO EXONS A4-
CC CC A5 OF THE COX1 GENE AND THUS MEDIATES THE INSERTION OF THE INTRON
CC CC INTO AN INTRONLESS STRAIN. RECOGNIZES AND CLEAVE THE SEQUENCE
CC CC 5'-TTTGATCTTGTGACCCCTGAAGTATA-3'.
CC CC -!- SIMILARITY: TO OTHER INTRON NUCLEASES.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; V00694; CAA24064.1; ALT_INIT.
CC CC PIR; A04511; QXBY34.
CC CC REBASE; 2616; I-SceII.
CC CC SGD; S0007264; A14.
CC CC InterPro; IPR001982; Endonuc_LAG/HNH.
CC CC Pfam; PF00961; LAGLIDADG_1; 2.
CC CC Mitochondrion; Hydrolase; Nuclease; Endonuclease; Intron homing.
CC CC KW SEQUENCE 235 AA; 27460 MW; 81A54BC755C2A131 CRC64;
CC CC
CC CC Query Match 31.2%; Score 5; DB 1; Length 235;
CC CC Best Local Similarity 100.0%; Pred. No. 63;
CC CC Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC QY 6 GIDFI 10
CC CC Db 97 GIDFI 101
CC CC
CC CC RESULT 22
CC CC SCOA_MYCTU STANDARD; PRT; 248 AA.
CC CC ID SCOA_MYCTU
CC CC AC 006167;
CC CC DT 01-NOV-1997 (Rel. 35, Created)
CC CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC CC DE Probable succinyl-CoA:3-ketoacid-coenzyme A transferase subunit A
CC CC DE (EC 2.8.3.5) (Succinyl CoA:3-oxoacid CoA-transferase) (OXCT A).
CC CC GN SCOA OR RV2504C OR MT2579 OR MTCY07A7.10C.
CC CC OS Mycobacterium tuberculosis.
CC CC OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
CC CC OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CC CC OX NCBI_TaxID=1773;
CC CC RN [1]
CC CC RP SEQUENCE FROM N.A.
CC CC RC STRAIN-H37RV;
CC CC RX MEDLINE=98295987; PubMed=9634230;
CC CC RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
CC CC RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekalia F.,
CC CC RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
CC CC RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
CC CC RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
CC CC RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
CC CC RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
CC CC RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
CC CC RT "Deciphering the biology of Mycobacterium tuberculosis from the
CC CC RT complete genome sequence.";
CC CC RL Nature 393:537-544(1998).
CC CC RN [2]
CC CC RP SEQUENCE FROM N.A.
CC CC RC STRAIN=CDC 1551 / Oshkosh;
CC CC RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
CC CC RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
CC CC RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
CC CC RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
CC CC RA Bishai W.;
CC CC RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
CC CC RT laboratory strains.";
CC CC RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
```

```
CC CC -!- CATALYTIC ACTIVITY: Succinyl-CoA + a 3-oxo acid - succinate + a 3-
CC CC oxo-acyl-CoA.
CC CC -!- SUBUNIT: HETERODIMER OF A SUBUNIT A AND A SUBUNIT B (BY
CC CC SIMILARITY).
CC CC -!- SIMILARITY: BELONGS TO THE 3-OXOACID COA-TRANSFERASE SUBUNIT A
CC CC FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; Z95556; CAB08922.1; -.
CC CC EMBL; AE007094; AAK46883.1; -.
CC CC TIGR; MT2579; -.
CC CC TuberculList; RV2504C; -.
CC CC InterPro; IPR004165; CoA_trans.
CC CC InterPro; IPR004163; CoA_transf_1.
CC CC Pfam; PF01144; CoA_trans; 1.
CC CC PROSITE; PS01273; CoA_TRANSF_1; 1.
CC CC Transferrase; Complete proteome.
CC CC FT DOMAIN 24 30 COA-BINDING (POTENTIAL).
CC CC SQ SEQUENCE 248 AA; 26276 MW; 81340193CC48EBD1 CRC64;
CC CC
CC CC Query Match 31.2%; Score 5; DB 1; Length 248;
CC CC Best Local Similarity 100.0%; Pred. No. 66;
CC CC Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC QY 4 NCGID 8
CC CC Db 53 NCGID 57
CC CC
CC CC RESULT 23
CC CC AOPA_HUMAN STANDARD; PRT; 264 AA.
CC CC ID AOPA_HUMAN
CC CC AC Q96PS8;
CC CC DT 15-JUN-2002 (Rel. 41, Created)
CC CC DT 15-JUN-2002 (Rel. 41, Last sequence update)
CC CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC CC DE Aquaporin 10 (Small intestine aquaporin).
CC CC GN AQP10.
CC CC OS Homo sapiens (Human).
CC CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC CC OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
CC CC OX NCBI_TaxID=9606;
CC CC RN [1]
CC CC RP SEQUENCE FROM N.A.
CC CC RC TISSUE=Jejunum;
CC CC RX MEDLINE=21458278; PubMed=11573934;
CC CC RA Hatakeyama S., Yoshida Y., Tani T., Koyama Y., Nihei K., Ohshiro K.,
CC CC RA Kamile J.-I., Yacita E., Suda T., Hatakeyama K., Yamamoto T.;
CC CC RT "Cloning of a new aquaporin (AQP10) abundantly expressed in duodenum
CC CC RT and jejunum.";
CC CC RL Biochem. Biophys. Res. Commun. 287:814-819(2001).
CC CC -!- FUNCTION: Forms a water channel. Not permeable to urea and
CC CC glycerol. May contribute to water transport in the upper portion
CC CC of small intestine.
CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC CC -!- TISSUE SPECIFICITY: Expressed exclusively in duodenum and jejunum.
CC CC Highest expression in absorptive epithelial cells at the tips of
CC CC villi in the jejunum.
CC CC -!- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
```

```
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF159174; AAL25998.1; -.
CC Genew; HGNC:16029; AQP10.
CC MIM; 606578; -.
CC InterPro; IPR000425; MIP_family.
CC Pfam; PF00230; MIP; 1.
CC ProDom; PD000295; MIP_family; 1.
CC PROSITE; PS00221; MIP; 1.
CC Transport; Transmembrane.
KW DOMAIN 1 27 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 28 48 POTENTIAL.
FT DOMAIN 49 52 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 53 73 POTENTIAL.
FT DOMAIN 74 101 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 102 122 POTENTIAL.
FT DOMAIN 123 156 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 157 177 POTENTIAL.
FT DOMAIN 178 187 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 188 208 POTENTIAL.
FT DOMAIN 209 264 EXTRACELLULAR (POTENTIAL).
FT SEQUENCE 264 AA; 27672 MW; 1332FF9533743412 CRC64;
SQ
Query Match 31.2%; Score 5; DB 1; Length 264;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 ANCGI 7
Db 207 ANCGI 211
RESULT 24
TRPA_PSEAE STANDARD; PRT; 268 AA.
ID TRPA_PSEAE
AC P07344;
DT 01-APR-1988 (Rel. 07, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tryptophan synthase alpha chain (EC 4.2.1.20).
GN TRPA OR PA0035.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=88174343; PubMed=3127651;
RA Hadero A., Crawford I.P.;
RT "Nucleotide sequence of the genes for tryptophan synthase in
Pseudomonas aeruginosa.";
RL Mol. Biol. Evol. 3:191-204(1986).
RN [2]
RP REVISIONS.
RA Crawford I.P., Eberly L.;
RL Submitted (MAR-1989) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of pseudomonas aeruginosa PA01, an
opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- FUNCTION: THE ALPHA SUBUNIT IS RESPONSIBLE FOR THE ALDOL CLEAVAGE
CC OF INDOLEGLYCEROL PHOSPHATE TO INDOLE AND GLYCERALDEHYDE 3-
CC PHOSPHATE.
CC
```

```
CC -!- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
CC = L-tryptophan + glyceraldehyde 3-phosphate + H(2)O.
CC -!- PATHWAY: Tryptophan biosynthesis; fifth (last) step.
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (by
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE TRPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M15826; AAA88463.1; -.
CC EMBL; AE004443; AAG03425.1; -.
CC PIR; B25355; TSPSAA.
CC HSP; P00929; 2WSY.
CC InterPro; IPR003009; FMN_enzyme.
CC InterPro; IPR002028; trp_synthaseA.
CC Pfam; PF00290; trp_syntA; 1.
CC ProDom; PD001535; Trp_synthaseA; 1.
CC TIGRfams; TIGR00262; trpA; 1.
CC PROSITE; PS00167; TRP_SYNTHASE_ALPHA; 1.
KW Tryptophan biosynthesis; Lyase; Complete proteome.
FT CONFLICT 42 73 A -> G (IN REF. 1).
FT CONFLICT 73 73 D -> E (IN REF. 1).
FT CONFLICT 148 148 I -> L (IN REF. 1).
FT SEQUENCE 268 AA; 28488 MW; 6EB7052524A9BEAD CRC64;
SQ
Query Match 31.2%; Score 5; DB 1; Length 268;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 GIDFI 10
Db 147 GIDFI 151
RESULT 25
TRPA_PSEPU STANDARD; PRT; 269 AA.
ID TRPA_PSEPU
AC P11081;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tryptophan synthase alpha chain (EC 4.2.1.20).
GN TRPA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PPG1 C1S;
RX MEDLINE=89335826; PubMed=2503057;
RA Eberly L., Crawford I.P.;
RT "DNA sequence of the tryptophan synthase genes of Pseudomonas
putida.";
RL Biochimie 71:521-531(1989).
CC -!- FUNCTION: THE ALPHA SUBUNIT IS RESPONSIBLE FOR THE ALDOL CLEAVAGE
CC OF INDOLEGLYCEROL PHOSPHATE TO INDOLE AND GLYCERALDEHYDE 3-
CC PHOSPHATE.
CC -!- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
CC = L-tryptophan + glyceraldehyde 3-phosphate + H(2)O.
CC -!- PATHWAY: Tryptophan biosynthesis; fifth (last) step.
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (by
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE TRPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
```



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X13299; CAA31662.1; -  
DR PIR; A30768; A30768.  
DR PIR; S03836; S03836.  
DR HSSP; P00929; 2WSY.  
DR InterPro; IPR003009; FMN\_enzyme.  
DR InterPro; IPR002028; Trp\_synthaseA.  
DR Pfam; PF00290; trp\_syntA; 1.  
DR ProDom; PD001535; Trp\_synthaseA; 1.  
DR TIGRFAMs; TIGR00262; trpA; 1.  
DR PROSITE; PS00167; TRP\_SYNTHASE\_ALPHA; 1.  
KW Tryptophan biosynthesis; Lyase.  
SQ SEQUENCE 269 AA; 28460 MW; 79FE339C965FD6F8 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 269;  
Best Local Similarity 100.0%; Pred.No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|||||  
Db 147 GIDFI 151

RESULT 26  
TRPA\_PSEY  
ID TRPA\_PSEY STANDARD; PRT; 269 AA.  
AC P34816;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Tryptophan synthase alpha chain (EC 4.2.1.20).  
GN TRPA.  
OS Pseudomonas syringae (pv. syringae).  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=321;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93138427; PubMed=8423001;  
RA Auerbach S., Gao J., Gussin G.N.;  
RT "Nucleotide sequences of the trpI, trpB, and trpA genes of  
RT pseudomonas syringae: positive control unique to fluorescent  
RT pseudomonads";  
RL Gene 123:25-32(1993).  
CC -!- FUNCTION: THE ALPHA SUBUNIT IS RESPONSIBLE FOR THE ALDOL CLEAVAGE  
CC OF INDOLEGLYCEROL PHOSPHATE TO INDOLE AND GLYCERALDEHYDE 3-  
CC PHOSPHATE.  
CC -!- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate  
CC = L-tryptophan + glyceraldehyde 3-phosphate + H(2)O.  
CC -!- PATHWAY: Tryptophan biosynthesis; fifth (last) step.  
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (by  
CC SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE TRPA FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M95710; AAA26013.1; -  
DR PIR; JQ2127; JQ2127.  
DR HSSP; P00929; 2WSY.  
DR InterPro; IPR003009; FMN\_enzyme.  
DR InterPro; IPR002028; Trp\_synthaseA.  
DR Pfam; PF00290; trp\_syntA; 1.

DR ProDom; PD001535; Trp\_synthaseA; 1.  
DR TIGRFAMs; TIGR00262; trpA; 1.  
DR PROSITE; PS00167; TRP\_SYNTHASE\_ALPHA; 1.  
KW Tryptophan biosynthesis; Lyase.  
SQ SEQUENCE 269 AA; 28439 MW; AD71A35FB49B3785 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 269;  
Best Local Similarity 100.0%; Pred.No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|||||  
Db 147 GIDFI 151

RESULT 27  
Y012\_MYCGE  
ID Y012\_MYCGE STANDARD; PRT; 287 AA.  
AC P47258;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein MG012.  
GN MG012.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=96026346; PubMed=7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RT "The minimal gene complement of Mycoplasma genitalium";  
RL Science 270:397-403(1995).  
CC -!- SIMILARITY: SOME, TO BACTERIAL RIBOSOMAL PROTEIN S6 MODIFICATION  
CC PROTEIN (RIMK).  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U39680; AAC71228.1; -  
DR TIGR; MG012; -  
DR InterPro; IPR004666; RimK\_fam.  
DR TIGRFAMs; TIGR00768; rimK\_fam; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 287 AA; 32733 MW; 7A14AF52A5AD7BBD CRC64;

Query Match 31.2%; Score 5; DB 1; Length 287;  
Best Local Similarity 100.0%; Pred.No. 73;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGIDF 9  
|||||  
Db 245 CGIDF 249

RESULT 28  
Y012\_MYCPN  
ID Y012\_MYCPN STANDARD; PRT; 288 AA.  
AC P75097;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)



DE Hypothetical protein MG012 homolog (D12\_orf288).

GN MPN016 OR MP138.

OS Mycoplasma pneumoniae.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI\_TaxID=2104;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 29342 / M129;

RX MEDLINE=97105885; PubMed=8948633;

RA Himmelreich R., Hilbert H., Plagens H., Pirk1 E., Li B.-C.,

RA Herrmann R.;

RT "Complete sequence analysis of the genome of the bacterium Mycoplasma

RT pneumoniae.";

RL Nucleic Acids Res. 24:4420-4449(1996).

CC -!- SIMILARITY: SOME, TO BACTERIAL RIBOSOMAL PROTEIN S6 MODIFICATION

CC PROTEIN (RIMK).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AE000016; AAB95786.1; -.

DR InterPro; IPR003135; ATP-grasp.

DR InterPro; IPR004666; RimK\_fam.

DR Pfam; PF02222; ATP-grasp; 1.

DR TIGRFAMs; TIGR00768; rimK\_fam; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 288 AA; 32435 MW; D0215C770C9BDF58 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 288;

Best Local Similarity 100.0%; Pred. No. 73;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGIDF 9

DB 245 CGIDF 249

RESULT 29

PPP4\_HUMAN

ID PPP4\_HUMAN STANDARD; PRT; 307 AA.

AC P33172;

DT 01-OCT-1993 (Rel. 27, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Serine/threonine protein phosphatase 4 (EC 3.1.3.16) (Pp4) (Protein

DE phosphatase X) (pp-X).

GN PPP4C OR PPP4 OR PPX.

OS Homo sapiens (Human), and

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606, 10090;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=Human;

RX MEDLINE=93129628; PubMed=1336397;

RA Brewis N.D., Cohen P.T.W.;

RT "Protein phosphatase X has been highly conserved during mammalian

RT evolution.";

RL Biochim. Biophys. Acta 1171:231-233(1992).

RN [2]

RP REVISION TO 75.

RA Cohen P.T.W.;

RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC SPECIES=Human, and Mouse;

RX MEDLINE=99057922; PubMed=9837938;

RA Hu M.C.-T., Tang-Oxley Q., Qiu W.R., Wang Y.-P.,

RA Mihindukulasuriya K.A., Afshar R., Tan T.-H.;

RT "Protein phosphatase X interacts with c-Rel and stimulates c-

RT Rel/nuclear factor kappaB activity.";

RL J. Biol. Chem. 273:33561-33565(1998).

RN [4]

RP SEQUENCE FROM N.A.

RC SPECIES=Human; TISSUE=Placenta;

RA Strausberg R.;

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RC SPECIES=Mouse; STRAIN=129/SvJ;

RX MEDLINE=21564173; PubMed=11707325;

RA Hu M.C.-T., Shui J.W., Mihindukulasuriya K.A., Tan T.-H.;

RT "Genomic structure of the mouse PP4 gene: a developmentally regulated

RT protein phosphatase.";

RL Gene 278:89-99(2001).

RN [6]

RP SEQUENCE FROM N.A.

RC SPECIES=Mouse; TISSUE=Breast tumor;

RA Strausberg R.;

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: COULD BE INVOLVED IN MICROTUBULE ORGANIZATION.

CC -!- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O -> a protein +

CC phosphate.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR; CENTROSOSES.

CC -!- SIMILARITY: BELONGS TO THE PPP FAMILY OF PHOSPHATASES. PP-X

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X70218; CAA49753.1; -.

DR EMBL; AF097996; AAC96318.1; -.

DR EMBL; AF088911; AAC96297.1; -.

DR EMBL; BC001416; AAH01416.1; -.

DR EMBL; AF378669; AAL35110.1; -.

DR EMBL; BC001993; AAH01993.1; -.

DR PIR; S28173; S28173.

DR HSSP; P08129; IFJM.

DR Genew; HGNC:9319; PPP4C.

DR MIM; 602035; -.

DR MGD; MGI:1891763; Ppp4c.

DR InterPro; IPR004843; M-ppetrase.

DR InterPro; IPR004844; S/T\_phosphatse.

DR Pfam; PF00149; Metallophos; 1.

DR PRINTS; PR00114; STPHPTASE.

DR ProDom; PD000252; S/T\_phosphatse; 1.

DR SMART; SM00156; PP2Ac; 1.

DR PROSITE; PS00125; SER\_THR\_PHOSPHATASE; 1.

KW Hydrolase; Iron; Manganese.

FT METAL 54 54 IRON (BY SIMILARITY).

FT METAL 56 56 IRON (BY SIMILARITY).

FT METAL 82 82 IRON AND MANGANESE (BY SIMILARITY).

FT METAL 114 114 MANGANESE (BY SIMILARITY).

FT ACT\_SITE 115 115 GENERAL ACID (BY SIMILARITY).

FT METAL 164 164 MANGANESE (BY SIMILARITY).

FT METAL 238 238 MANGANESE (BY SIMILARITY).

SQ SEQUENCE 307 AA; 35080 MW; D6FE470A5C6CBCAC CRC64;

Query Match 31.2%; Score 5; DB 1; Length 307;

Best Local Similarity 100.0%; Pred. No. 77;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12

DB 282 DFIIIF 286

```
RESULT 30
PPP4_RABIT
ID PPP4_RABIT STANDARD; PRT; 307 AA.
AC P11084;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Serine/threonine protein phosphatase 4 (EC 3.1.3.16) (Pp4) (Protein
DE phosphatase X) (pp-X).
GN PPP4C.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Liver;
RX MEDLINE=90346193; PubMed=2166691;
RA Cohen P.T.W., Brewis N.D., Hughes V., Mann D.J.;
RT "Protein serine/threonine phosphatases; an expanding family.";
RL FEBS Lett. 268:355-359(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Liver;
RX MEDLINE=93209245; PubMed=8384557;
RA Brewis N.D., Street A.J., Prescott A.R., Cohen P.T.W.;
RT "PPX, a novel protein serine/threonine phosphatase localized to
RT centrosomes.";
RL EMBO J. 12:987-996(1993).
RN [3]
RP SEQUENCE OF 105-307 FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Liver;
RX MEDLINE=89078593; PubMed=2849555;
RA da Cruz e Silva O.B., da Cruz e Silva E.F., Cohen P.T.W.;
RT "Identification of a novel protein phosphatase catalytic subunit by
RT cDNA cloning.";
RL FEBS Lett. 242:106-110(1988).
CC -1- FUNCTION: COULD BE INVOLVED IN MICROTUBULE ORGANIZATION.
CC -1- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O - a protein +
CC phosphate.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR; CENTROSOMES.
CC -1- SIMILARITY: BELONGS TO THE PPP FAMILY OF PHOSPHATASES. PP-X
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14031; CAA32191.1; -.
DR EMBL; S57412; AAB25913.1; -.
DR PIR; S11059; PARBA2.
DR PIR; S02006; S02006.
DR HSSP; P08129; 1FJM.
DR InterPro; IPR004843; M-pppestrase.
DR InterPro; IPR004844; S/T_phosphatse.
DR Pfam; PF00149; Metallophos; 1.
DR PRINTS; PR00114; STPHPTASE.
DR ProDom; PD000252; S/T_phosphatse; 1.
DR SMART; SM00156; PP2Ac; 1.
DR PROSITE; PS00125; SER_THR_PHOSPHATASE; 1.
KW Hydrolase; Iron; Manganese.
FT METAL 54 54 IRON (BY SIMILARITY).
FT METAL 56 56 IRON (BY SIMILARITY).
FT METAL 82 82 IRON AND MANGANESE (BY SIMILARITY).
FT METAL 114 114 MANGANESE (BY SIMILARITY).
FT ACT_SITE 115 115 GENERAL ACID (BY SIMILARITY).
FT METAL 164 164 MANGANESE (BY SIMILARITY).
```

```
FT METAL 238 238 MANGANESE (BY SIMILARITY).
SQ SEQUENCE 307 AA; 35037 MW; 364A1641F8B22B41 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 307;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12
Db 282 DFIIF 286

RESULT 31
YSI9_STRCO
ID YSI9_STRCO STANDARD; PRT; 310 AA.
AC Q9ZBI5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein SCO6489.
GN SCO6489 OR SC9C7.25.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Harris K.D., James K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris K.F., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -1- SIMILARITY: BELONGS TO THE LDCA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL035161; CAA22737.1; -.
DR InterPro; IPR003507; UPF0094.
DR Pfam; PF02016; UPF0094; 1.
KW Hypothetical protein; Hydrolase; Complete proteome.
SQ SEQUENCE 310 AA; 32763 MW; 6E2C587F3AE8E615 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 310;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10
Db 141 GIDFI 145

RESULT 32
YC57_ARCFU
ID YC57_ARCFU STANDARD; PRT; 320 AA.
AC O29011;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
```

DE Hypothetical protein AF1257.  
GN AF1257.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,  
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 390:364-370(1997).  
CC -!- SIMILARITY: BELONGS TO THE UPF0020 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF01257; AAB89985.1; -.  
DR TIGR; AF1257; -.  
DR InterPro; IPR000241; RNA\_methylase.  
DR InterPro; IPR000051; SAM\_bind.  
DR InterPro; IPR004114; THUMP\_dom.  
DR Pfam; PF01170; UPF0020; 1.  
DR Pfam; PF02926; THUMP; 1.  
DR TIGRFAMS; TIGR01177; TIGR01177; 1.  
DR PROSITE; PS01261; UPF0020; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 320 AA; 36673 MW; A879F320AA8CD637 CRC64;  
Query Match 31.2%; Score 5; DB 1; Length 320;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 GIDFI 10  
Db 203 GIDFI 207  
RESULT 33  
MRAY\_ENTFA  
ID MRAY\_ENTFA STANDARD; PRT; 321 AA.  
AC O07107;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-  
MURNAc-pentapeptide phosphotransferase).  
GN MRAY.  
OS Enterococcus faecalis (Streptococcus faecalis).  
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
OX NCBI\_TaxID=1351;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A24836;  
RX MEDLINE=97431524; PubMed=9287029;  
RA Pucci M.J., Thanassi J.A., Discotto L.F., Kessler R.E.,

RA Dougherty T.J.;  
RT "Identification and characterization of cell wall-cell division gene  
clusters in pathogenic Gram-positive cocci.";  
RL J. Bacteriol. 179:5632-5635(1997).  
CC -!- FUNCTION: FIRST STEP OF THE CELL WALL PEPTIDOGLYCAN.  
CC BIOSYNTHESIS OF THE CELL WALL PEPTIDOGLYCAN.  
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-L-  
lysyl-D-alanyl-D-alanine + undecaprenyl phosphate = UMP + N-  
acetylmuramoyl-L-alanyl-D-glutamyl-L-lysyl-D-alanyl-D-alanine-  
diphosphoundecaprenol.  
CC -!- PATHWAY: Peptidoglycan biosynthesis.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 4. MRAY  
CC SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U94707; AAC45634.1; -.  
DR InterPro; IPR000715; Glycos\_transf\_4.  
DR InterPro; IPR003524; PNacPP\_transf.  
DR Pfam; PF00953; Glycos\_transf\_4; 1.  
DR TIGRFAMS; TIGR00445; mray; 1.  
DR PROSITE; PS01347; MRAY\_1; 1.  
DR PROSITE; PS01348; MRAY\_2; 1.  
KW Peptidoglycan synthesis; Cell division; Transferase; Transmembrane.  
FT TRANSMEM 6 26 POTENTIAL.  
FT TRANSMEM 54 74 POTENTIAL.  
FT TRANSMEM 77 97 POTENTIAL.  
FT TRANSMEM 117 137 POTENTIAL.  
FT TRANSMEM 143 163 POTENTIAL.  
FT TRANSMEM 175 195 POTENTIAL.  
FT TRANSMEM 200 220 POTENTIAL.  
FT TRANSMEM 239 259 POTENTIAL.  
FT TRANSMEM 301 321 POTENTIAL.  
SQ SEQUENCE 321 AA; 35834 MW; EB4FE3283C29C344 CRC64;  
Query Match 31.2%; Score 5; DB 1; Length 321;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 FIIFW 13  
Db 157 FIIFW 161  
RESULT 34  
MCE\_CHVP1  
ID MCE\_CHVP1 STANDARD; PRT; 330 AA.  
AC Q84424;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE mRNA capping enzyme (mRNA guanylyltransferase) (EC 2.7.7.50)  
DE (GTP--RNA guanylyltransferase).  
GN A103R.  
OS Paramesidium bursaria chlorella virus 1 (PBCV-1).  
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.  
OX NCBI\_TaxID=10506;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95407089; PubMed=7676624;  
RA Li Y., Lu Z., Burbank D.E., Kutish G.F., Rock D.L., Etten J.L.;  
RT "Analysis of 43 kb of the Chlorella virus PBCV-1 330-kb genome: map  
positions 45 to 88.";  
RL Virology 212:134-150(1995).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).

RX MEDLINE=97304383; PubMed=9160746;  
RA Haakansson K., Doherty A.J., Shuman S., Wigley D.B.;  
RT "X-ray crystallography reveals a large conformational change during  
RL guanyl transfer by mRNA capping enzymes.";  
RN Cell 89:545-553(1997).  
RP [3]  
RX X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 11-327.  
RA MEDLINE=98132620; PubMed=9465045;  
RX Haakansson K., Wigley D.B.;  
RA "Structure of a complex between a cap analogue and mRNA guanylyl  
RT transferase demonstrates the structural chemistry of RNA capping.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:1505-1510(1998).  
RN [-] FUNCTION: M-RNA CAPPING. TRANSFERS A GMP CAP ONTO THE END OF MRNA  
CC THAT TERMINATES WITH A 5'-DIPHOSPHATE TAIL.  
CC [-] CATALYTIC ACTIVITY: GTP + (5')PP-pur-mRNA - diphosphate +  
CC G(5')PPP-pur-mRNA.  
CC [-] COFACTOR: MAGNESIUM OR MANGANESE.  
CC [-] SUBUNIT: MONOMER.  
CC [-] SIMILARITY: BELONGS TO THE EUKARYOTIC GTASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U42580; AAC96471.1; -.  
DR PDB; 1CKM; 07-JUL-97.  
DR PDB; 1CKN; 07-JUL-97.  
DR PDB; 1CKO; 28-JAN-98.  
DR InterPro; IPR001339; mRNA\_cap\_enzyme.  
DR Pfam; PF01331; mRNA\_cap\_enzyme; 1.  
KW transferase; Nucleotidyltransferase; mRNA processing; mRNA capping;  
KW 3D-structure.  
FT ACT\_SITE 82 82 GUANYLYLATION SITE.  
SQ SEQUENCE 330 AA; 37832 MW; 6AF8A404710812D9 CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 330;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 7 IDP11 11  
Db 243 IDP11 247  
  
RESULT 35  
AT9B\_HUMAN STANDARD; PRT; 337 AA.  
ID AT9B\_HUMAN  
AC O43861; O60872;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Potential phospholipid-transporting ATPase IIB (EC 3.6.3.1) (HUSSEY-20)  
DE (Fragment).  
GN ATP9B OR ATP11B OR NEOL1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=98217376; PubMed=9548971;  
RA Halleck M.S., Pradhan D., Blackman C.F., Berkes C., Williamson P.L.,  
RA Schlegel R.A.;  
RT "Multiple members of a third subfamily of p-type ATPases identified by  
RT genomic sequences and ESTs.";  
RL Genome Res. 8:354-361(1998).  
RN [2]  
RP SEQUENCE OF 10-337 FROM N.A.

RC TISSUE=Brain;  
RX MEDLINE=21064499; PubMed=11124703;  
RA Stanchi F., Bertocco E., Toppo S., Dioguardi R., Simionati B.,  
RA Cannata N., Zimbello R., Lanfranchi G., Valle G.;  
RT "Characterization of 16 novel human genes showing high similarity to  
RT yeast sequences";  
RL Yeast 18:69-80(2001).  
CC [-] CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.  
CC [-] SUBCELLULAR LOCATION: Integral membrane protein.  
CC [-] SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2  
CC ATPASES). SUBFAMILY IV.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U78978; AAC05243.1; -.  
DR EMBL; AJ006268; CAA06934.1; -.  
DR Genew; HGNC:13541; ATP9B.  
DR InterPro; IPR001757; ATPase\_E1-E2.  
DR PROSITE; PS00154; ATPASE\_E1\_E2; PARTIAL.  
KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;  
KW Multigene family.  
FT NON\_TER 1 1  
FT DOMAIN 1 120 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 121 141 POTENTIAL.  
FT DOMAIN 142 153 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 154 172 POTENTIAL.  
FT DOMAIN 173 202 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 203 221 POTENTIAL.  
FT DOMAIN 222 228 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 229 251 POTENTIAL.  
FT DOMAIN 252 257 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 258 278 POTENTIAL.  
FT DOMAIN 279 295 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 296 320 POTENTIAL.  
FT DOMAIN 321 337 CYTOPLASMIC (POTENTIAL).  
FT METAL 64 64 MAGNESIUM (BY SIMILARITY).  
FT METAL 68 68 MAGNESIUM (BY SIMILARITY).  
FT CONFLICT 76 76 N -> D (IN REF. 2).  
FT CONFLICT 196 196 K -> R (IN REF. 2).  
FT CONFLICT 210 210 I -> S (IN REF. 2).  
FT CONFLICT 226 226 D -> E (IN REF. 2).  
FT CONFLICT 263 263 D -> E (IN REF. 2).  
FT CONFLICT 285 285 K -> R (IN REF. 2).  
FT CONFLICT 293 293 N -> D (IN REF. 2).  
FT CONFLICT 321 321 N -> K (IN REF. 2).  
SQ SEQUENCE 337 AA; 37498 MW; E93C93A44BD826B2 CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 337;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 ANCGI 7  
Db 75 ANCGI 79  
  
RESULT 36  
YHJD\_ECOLI STANDARD; PRT; 337 AA.  
ID YHJD\_ECOLI  
AC P37642;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein yhjD.  
GN YHJD OR B3522.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;



OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94316500; PubMed=8041620;  
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
RT region from 76.0 to 81.5 minutes.";  
RL Nucleic Acids Res. 22:2576-2586(1994).  
CC -|- SIMILARITY: STRONG, TO E.CHRYSANTHEMI HYPOTHETICAL PROTEIN IN  
CC KDGK 5'REGION (AC P45417).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U00039; AAB18498.1; -.  
DR EMBL; AE000428; AAC76547.1; -.  
DR EcoGene; EG12248; yhjD.  
DR InterPro; IPR005274; Cons\_hypoth766.  
DR InterPro; IPR004664; RNase\_BN.  
DR Pfam; PF03631; Ribonuclease\_BN; 1.  
DR TIGRFAMS; TIGR00766; TIGR00766; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 337 AA; 37911 MW; C41B2A224902E311 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 337;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 FWIFW 16  
|||||  
Db 246 FWIFW 250

RESULT 37  
SYFA\_AQUAE  
ID SYFA\_AQUAE STANDARD; PRT; 338 AA.  
AC O67087;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Phenylalanyl-tRNA synthetase alpha chain (EC 6.1.1.20) (Phenylalanine-  
DE -tRNA ligase alpha chain) (PHERS).  
GN PHER OR AQ\_953.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;  
OC Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
RT aeolicus";  
RL Nature 392:353-358(1998).  
CC -|- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +  
CC diphosphate + L-phenylalanyl-tRNA(Phe).  
CC -|- SUBUNIT: Tetramer of two alpha and two beta chains (BY  
CC SIMILARITY).  
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -|- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC PHE-TRNA SYNTHETASE ALPHA CHAIN SUBFAMILY 1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AE000715; AAC07051.1; -.  
DR HSSP; P27001; 1PYS.  
DR InterPro; IPR002106; AATRNA\_ligaseII.  
DR InterPro; IPR004529; PheS.  
DR InterPro; IPR004188; Phe\_tRNA\_synt\_N.  
DR InterPro; IPR002319; tRNA\_synt\_2d.  
DR Pfam; PF01409; tRNA\_synt\_2d; 1.  
DR Pfam; PF02912; Phe\_tRNA\_synt\_N; 1.  
DR TIGRFAMS; TIGR00468; PheS; 1.  
DR PROSITE; PS00862; AA\_TRNA\_LIGASE\_II; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
SQ SEQUENCE 338 AA; 39504 MW; DC36592C22FB2305 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 338;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
|||||  
Db 293 NCGID 297

RESULT 38  
CYS1\_OSTOS  
ID CYS1\_OSTOS STANDARD; PRT; 341 AA.  
AC P25802;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cathepsin B-like cysteine proteinase 1 precursor (EC 3.4.22.-).  
GN CP-1.  
OS Ostertagia ostertagi.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
OC Trichostrongyloidea; Haemonchidae; Ostertagiinae; Ostertagia.  
OX NCBI\_TaxID=6317;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Larva;  
RX MEDLINE=93116804; PubMed=1475000;  
RA Pratt D., Boisvenue R.J., Cox G.N.;  
RT "Isolation of putative cysteine protease genes of Ostertagia  
RT ostertagi";  
RL Mol. Biochem. Parasitol. 56:39-48(1992).  
CC -|- FUNCTION: EXPRESSION OF THE PROTEASE CORRELATES WITH BLOOD-FEEDING  
CC AND SUGGESTS A ROLE FOR THE PROTEASE IN BLOOD DIGESTION.  
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M88503; AAA29433.1; -.  
DR EMBL; M88503; AAA29434.1; ALT\_SEQ.  
DR EMBL; M88504; AAA29435.1; -.  
DR HSSP; P07688; IQDQ.  
DR MEROPS; C01.101; -.  
DR InterPro; IPR000668; Peptidase\_C1.  
DR InterPro; IPR000169; SHprot\_acsite.  
DR Pfam; PF00112; Peptidase\_C1; 1.  
DR PRINTS; PR00705; PAPAIN.  
DR proDom; PD000158; Peptidase\_C1; 1.

DR PROSITE; PS00139; THIOLEPROTEASE\_CYS; 1.  
DR PROSITE; PS00639; THIOLEPROTEASE\_HIS; 1.  
DR PROSITE; PS00640; THIOLEPROTEASE\_HIS; 1.  
KW Hydrolase; Thiol protease; Zymogen; Glycoprotein; Multigene family;  
KW Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT PROPEP 20 88 ACTIVATION PEPTIDE (POTENTIAL).  
FT CHAIN 89 341 CATHEPSIN B-LIKE CYSTEINE PROTEINASE 1.  
FT ACT\_SITE 119 119 BY SIMILARITY.  
FT ACT\_SITE 288 288 BY SIMILARITY.  
FT ACT\_SITE 308 308 BY SIMILARITY.  
FT DISULFID 104 133 BY SIMILARITY.  
FT DISULFID 116 160 BY SIMILARITY.  
FT DISULFID 152 218 BY SIMILARITY.  
FT DISULFID 153 156 BY SIMILARITY.  
FT DISULFID 189 222 BY SIMILARITY.  
FT DISULFID 197 209 BY SIMILARITY.  
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 341 AA; 38439 MW; 07968646E3D920F6 CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 341;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 QANCG 6  
Db 113 QANCG 117  
  
RESULT 39  
CYS1\_HAECO STANDARD; PRT; 342 AA.  
ID CYS1\_HAECO  
AC P19092;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cathepsin B-like cysteine proteinase 1 precursor (EC 3.4.22.-).  
GN AC-1.  
OS Haemochus contortus (Barber pole worm).  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.  
OX NCBI\_TaxID=6289;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Isolate BPL1;  
RX MEDLINE=90348715; PubMed=2385265;  
RA Cox G.N., Pratt D., Hageman R., Boisvenue R.J.;  
RT "Molecular cloning and primary sequence of a cysteine protease  
expressed by Haemochus contortus adult worms.";  
RL Mol. Biochem. Parasitol. 41:25-34(1990).  
CC -!- FUNCTION: EXPRESSION OF THE PROTEASE CORRELATES WITH BLOOD-FEEDING  
CC -!- AND SUGGESTS A ROLE FOR THE PROTEASE IN BLOOD DIGESTION.  
CC -!- DEVELOPMENTAL STAGE: AT LOW LEVEL IN THE THIRD AND FOURTH-STAGE  
CC -!- LARVAE, AND ABUNDANT IN ADULT WORMS.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M31112; AAA29175.1; -.  
DR PIR; A45524; A45524.  
DR HSSP; P07688; 1QDQ.  
DR MEROPS; C01.101; -.  
DR InterPro; IPR000668; Peptidase\_C1.  
DR InterPro; IPR000169; SHprot\_acsite.  
DR Pfam; PF00112; Peptidase\_C1; 1.  
DR PRINTS; PR00705; PAPA1N.

DR ProDom; PD000158; Peptidase\_C1; 1.  
DR PROSITE; PS00139; THIOLEPROTEASE\_CYS; 1.  
DR PROSITE; PS00639; THIOLEPROTEASE\_HIS; 1.  
DR PROSITE; PS00640; THIOLEPROTEASE\_HIS; 1.  
KW Hydrolase; Thiol protease; Zymogen; Glycoprotein; Signal;  
KW Multigene family.  
FT SIGNAL 1 18 POTENTIAL.  
FT PROPEP 19 86 ACTIVATION PEPTIDE (POTENTIAL).  
FT CHAIN 87 342 CATHEPSIN B-LIKE CYSTEINE PROTEINASE 1.  
FT ACT\_SITE 114 114 BY SIMILARITY.  
FT ACT\_SITE 285 285 BY SIMILARITY.  
FT ACT\_SITE 305 305 BY SIMILARITY.  
FT DISULFID 100 128 BY SIMILARITY.  
FT DISULFID 111 156 BY SIMILARITY.  
FT DISULFID 147 214 BY SIMILARITY.  
FT DISULFID 148 152 BY SIMILARITY.  
FT DISULFID 185 218 BY SIMILARITY.  
FT DISULFID 193 205 BY SIMILARITY.  
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 342 AA; 38459 MW; D33D62F7419F0471 CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 342;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 QANCG 6  
Db 108 QANCG 112  
  
RESULT 40  
CYS2\_HAECO STANDARD; PRT; 342 AA.  
ID CYS2\_HAECO  
AC P25793;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cathepsin B-like cysteine proteinase 2 precursor (EC 3.4.22.-).  
GN AC-2.  
OS Haemochus contortus (Barber pole worm).  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.  
OX NCBI\_TaxID=6289;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91218800; PubMed=2090940;  
RA Pratt D., Cox G.N., Milhausen M.J., Boisvenue R.J.;  
RT "A developmentally regulated cysteine protease gene family in  
Haemochus contortus.";  
RL Mol. Biochem. Parasitol. 43:181-192(1990).  
CC -!- FUNCTION: EXPRESSION OF THE PROTEASE CORRELATES WITH BLOOD-FEEDING  
CC -!- AND SUGGESTS A ROLE FOR THE PROTEASE IN BLOOD DIGESTION.  
CC -!- DEVELOPMENTAL STAGE: AT LOW LEVEL IN THE THIRD AND FOURTH-STAGE  
CC -!- LARVAE, AND ABUNDANT IN ADULT WORMS.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M60213; AAA29171.1; -.  
DR EMBL; M60212; AAA29171.1; JOINED.  
DR PIR; A44965; A44965.  
DR HSSP; P07688; 1QDQ.  
DR MEROPS; C01.101; -.  
DR InterPro; IPR000668; Peptidase\_C1.





```
FT  TRANSMEM  175  195  POTENTIAL.
FT  TRANSMEM  207  227  POTENTIAL.
FT  TRANSMEM  242  262  POTENTIAL.
FT  TRANSMEM  276  296  POTENTIAL.
FT  TRANSMEM  304  324  POTENTIAL.
SQ  SEQUENCE  342 AA; 39534 MW; 08EFEC3E2C4955D8 CRC64;

Query Match
Best Local Similarity 31.2%; Score 5; DB 1; Length 342;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  6 GIDFI 10
Db  93 GIDFI 97

RESULT 43
AROB_HELPJ
ID  AROB_HELPJ  STANDARD;  PRT;  343 AA.
AC  Q92MF2;
DT  30-MAY-2000 (Rel. 39, Created)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DE  15-JUN-2002 (Rel. 41, Last annotation update)
DE  3-dehydroquininate synthase (EC 4.2.3.4).
GN  AROB OR JHP0268.
OS  Helicobacter pylori J99 (Campylobacter pylori J99).
OC  Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC  Helicobacter.
OX  NCBI_TaxID=85963;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=99120557; PubMed=9923682;
RA  Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA  Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA  Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA  Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA  Trust T.J.;
RT  "Genomic sequence comparison of two unrelated isolates of the human
RT  gastric pathogen Helicobacter pylori."
RL  Nature 397:176-180(1999).
CC  -1- CATALYTIC ACTIVITY: 3-deoxy-arabino-heptulosonate 7-phosphate - 3-
CC  dehydroquininate + phosphate.
CC  -1- COFACTOR: NAD and a divalent metal cation (By similarity).
CC  -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC  second step.
CC  -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC  -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC  -1- SIMILARITY: BELONGS TO THE DEHYDROQUINATE SYNTHASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE001464; AAD05849.1; -
DR  HSSP; P07547; 1DQS.
DR  InterPro; IPR002658; DHQ_synthase.
DR  Pfam; PF01761; DHQ_synthase; 1.
KW  Aromatic amino acid biosynthesis; Lyase; NAD; Complete proteome.
SQ  SEQUENCE 343 AA; 39106 MW; 21BDA1167236ED5B CRC64;

Query Match
Best Local Similarity 31.2%; Score 5; DB 1; Length 343;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  6 GIDFI 10
Db  111 GIDFI 115
```

```
RESULT 44
AROB_HELPY
ID  AROB_HELPY  STANDARD;  PRT;  343 AA.
AC  P56081;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  3-dehydroquininate synthase (EC 4.2.3.4).
GN  AROB OR HP0283.
OS  Helicobacter pylori (Campylobacter pylori).
OC  Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC  Helicobacter.
OX  NCBI_TaxID=210;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=26695 / ATCC 700392;
RX  MEDLINE=97394467; PubMed=9252185;
RA  Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA  Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA  Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA  Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA  McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA  Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA  Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA  Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA  Venter J.C.;
RT  "The complete genome sequence of the gastric pathogen Helicobacter
RT  pylori."
RL  Nature 388:539-547(1997).
CC  -1- CATALYTIC ACTIVITY: 3-deoxy-arabino-heptulosonate 7-phosphate - 3-
CC  dehydroquininate + phosphate.
CC  -1- COFACTOR: NAD and a divalent metal cation (By similarity).
CC  -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC  second step.
CC  -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC  -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC  -1- SIMILARITY: BELONGS TO THE DEHYDROQUINATE SYNTHASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE000547; AAD07351.1; -
DR  HSSP; P07547; 1DQS.
DR  TIGR; HP0283; -
DR  InterPro; IPR002658; DHQ_synthase.
DR  Pfam; PF01761; DHQ_synthase; 1.
KW  Aromatic amino acid biosynthesis; Lyase; NAD; Complete proteome.
SQ  SEQUENCE 343 AA; 39120 MW; 5AEAC2F4DE816D13 CRC64;

Query Match
Best Local Similarity 31.2%; Score 5; DB 1; Length 343;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  6 GIDFI 10
Db  111 GIDFI 115

RESULT 45
AROB_CAMJE
ID  AROB_CAMJE  STANDARD;  PRT;  351 AA.
AC  Q9PNT2;
DT  15-JUN-2002 (Rel. 41, Created)
DT  15-JUN-2002 (Rel. 41, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  3-dehydroquininate synthase (EC 4.2.3.4).
GN  AROB OR CJ1008C.
OS  Campylobacter jejuni.
```



OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
OC Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NTC 11168;  
RX MEDLINE=20150912; PubMed=10688204;  
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Barrell B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
RT reveals hypervariable sequences.";  
RL Nature 403:665-668(2000).  
CC -|- CATALYTIC ACTIVITY: 3-deoxy-arabino-heptulosonate 7-phosphate = 3-  
CC dehydroquinate + phosphate.  
CC -|- COFACTOR: NAD and a divalent metal cation (By similarity).  
CC -|- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;  
CC second step.  
CC -|- SUBUNIT: Monomer (By similarity).  
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -|- SIMILARITY: BELONGS TO THE DEHYDROQUINATE SYNTHASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AL139076; CAB73264.1; -.  
DR HSP; P07547; 1DQS.  
DR InterPro; IPR002658; DHQ\_synthase.  
DR Pfam; PF01761; DHQ\_synthase; 1.  
KW Aromatic amino acid biosynthesis; Lyase; NAD; Complete proteome.  
SQ SEQUENCE 351 AA; 39596 MW; DEF5FF894DE8727 CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 351;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 6 GIDFI 10  
Db 110 GIDFI 114

Search completed: May 11, 2003, 20:12:49  
Job time : 5.25664 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 19:33:41 ; Search time 8.49557 Seconds  
(without alignments)  
388.055 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	43.8	483	8 Q36097	Q36097 theileria p
2	6	37.5	156	17 O27080	O27080 methanobact
3	6	37.5	205	16 O25493	O25493 helicobacte
4	6	37.5	250	16 Q8REQ8	Q8req8 fusobacteri
5	6	37.5	319	5 Q9XYQ1	Q9xyqi entamoeba h
6	6	37.5	328	5 Q9N2T4	Q9n2t4 caenorhabdi
7	6	37.5	330	5 Q23072	Q23072 caenorhabdi
8	6	37.5	360	16 O87850	O87850 streptomyce
9	6	37.5	399	12 Q68409	Q68409 human cytom
10	6	37.5	430	10 Q9SMW7	Q9smw7 arabidopsis
11	6	37.5	674	12 Q9WI36	Q9wi36 cauliflower
12	6	37.5	680	12 Q83169	Q83169 cauliflower
13	6	37.5	680	12 Q66162	Q66162 cauliflower
14	6	37.5	1091	16 Q9KAQ7	Q9kaq7 bacillus ha
15	6	37.5	1285	12 Q8V3M7	Q8v3m7 swinepox vi
16	5	31.2	41	16 Q9PIF6	Q9pif6 campylobact

17	5	31.2	52	8 Q9ZZU6	Q9zzu6 littorina s
18	5	31.2	57	16 Q8RH01	Q8rh01 fusobacteri
19	5	31.2	83	2 Q9AIF7	Q9aif7 carsonella
20	5	31.2	90	17 Q8TMA6	Q8tma6 methanosarc
21	5	31.2	103	2 Q9Z5P6	Q9z5p6 enterococcu
22	5	31.2	106	16 Q55541	Q55541 synechocyst
23	5	31.2	107	10 Q9S6Y4	Q9s6y4 lotus corni
24	5	31.2	107	10 Q9S6Y3	Q9s6y3 lotus corni
25	5	31.2	108	2 Q936F3	Q936f3 staphylococ
26	5	31.2	113	2 Q8VVE1	Q8vve1 uncultured
27	5	31.2	118	2 Q9FDG9	Q9fdg9 streptococc
28	5	31.2	120	16 Q8Y6N6	Q8y6n6 listeria mo
29	5	31.2	123	2 Q68605	Q68605 campylobact
30	5	31.2	126	4 Q95545	Q95545 homo sapien
31	5	31.2	128	2 Q53372	Q53372 synechocyst
32	5	31.2	128	17 Q26942	Q26942 methanobact
33	5	31.2	134	17 Q9HK39	Q9hk39 thermoplasm
34	5	31.2	136	4 Q13964	Q13964 homo sapien
35	5	31.2	136	5 Q21953	Q21953 caenorhabdi
36	5	31.2	138	16 Q97KA8	Q97ka8 clostridium
37	5	31.2	147	17 Q974Q5	Q974q5 sulfolobus
38	5	31.2	148	4 Q9NWH4	Q9nwh4 homo sapien
39	5	31.2	155	16 Q8REG9	Q8reg9 fusobacteri
40	5	31.2	158	2 Q9RFA1	Q9rfal salmonella
41	5	31.2	160	16 Q97L62	Q97l62 clostridium
42	5	31.2	161	16 Q8XKN7	Q8xkn7 clostridium
43	5	31.2	163	16 Q8YPA6	Q8ypa6 anabaena sp
44	5	31.2	173	16 Q99TW3	Q99uw3 staphylococ
45	5	31.2	178	4 Q8WXJ2	Q8wxj2 homo sapien
46	5	31.2	180	16 Q8U6F5	Q8u6f5 agrobacteri
47	5	31.2	181	16 Q9PQL6	Q9pql6 ureaplasma
48	5	31.2	187	16 Q9PIA4	Q9pia4 campylobact
49	5	31.2	197	5 Q9GPR4	Q9gpr4 dictyosteli
50	5	31.2	208	16 Q927E8	Q927e8 listeria in
51	5	31.2	208	16 Q8Y3Y6	Q8y3y6 listeria mo
52	5	31.2	211	4 Q9NRL8	Q9nrl8 homo sapien
53	5	31.2	212	5 Q95VB9	Q95vb9 tetrahymena
54	5	31.2	214	5 Q20964	Q20964 caenorhabdi
55	5	31.2	214	5 Q20968	Q20968 caenorhabdi
56	5	31.2	214	14 P97966	P97966 unidentified
57	5	31.2	214	16 Q8RD62	Q8rd62 thermoanaer
58	5	31.2	215	11 P97470	P97470 mus musculu
59	5	31.2	216	6 Q9N224	Q9n224 gorilla gor
60	5	31.2	216	6 Q9N222	Q9n222 gorilla gor
61	5	31.2	216	6 Q9N1Z7	Q9nlz7 pongo pygma
62	5	31.2	217	2 Q8RR39	Q8rr39 streptococc
63	5	31.2	217	10 Q8S1Q1	Q8slq1 oryza sativ
64	5	31.2	219	16 Q9JQX9	Q9jqx9 neisseria m
65	5	31.2	220	16 Q8XMG7	Q8xmg7 clostridium

ALIGNMENTS

RESULT 1				
Q36097				
ID	Q36097	PRELIMINARY;	PRT;	483 AA.
AC	Q36097;			
DT	01-NOV-1996 (TReMBLrel. 01, Created)			
DT	01-NOV-1996 (TReMBLrel. 01, Last sequence update)			
DT	01-JUN-2002 (TReMBLrel. 21, Last annotation update)			
DE	Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).			
GN	COI.			
OS	Theileria parva.			
OG	Mitochondrion.			
OC	Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;			
OC	Theileria.			
OX	NCBI_TaxID=5875;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MUGUGA;			
RX	MEDLINE=94155854; Pubmed=8112303;			
RA	Kairo A., Fairlamb A., Gobright E., Nene V.;			

RT "A 7.1 kb linear DNA molecule of Theileria parva has scrambled rDNA  
 RT sequences and open reading frames for mitochondrially-encoded  
 RT proteins.";  
 RL EMBO J. 13:898-905(1994).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC 3 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 CC AND COPPER B (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME  
 CC C + 2 H(2)O.  
 CC -1- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 DR EMBL; Z23263; CAA80798.1; -.  
 DR InterPro; IPR000883; COX1.  
 DR Pfam; PF00115; COX1; 1.  
 KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
 KW Respiratory chain; Transmembrane.  
 FT NON\_TER 1  
 SQ SEQUENCE 483 AA; 54008 MW; 314438D6EF4CF3D6 CRC64;

Query Match 43.8%; Score 7; DB 8; Length 483;  
 Best Local Similarity 100.0%; Pred. No. 3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 12

|||||

Db 150 GIDFII 156

## RESULT 2

O27080 ID O27080 PRELIMINARY; PRT; 156 AA.  
 AC O27080;

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE N-terminal acetyltransferase complex, subunit ARD1.  
 GN MTH999.

OS Methanobacterium thermoautotrophicum.  
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
 OC Methanobacteriaceae; Methanothermobacter.  
 OX NCBI\_TaxID=187420;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DELTA H;

RX MEDLINE=98037514; PubMed=9371463;

RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,  
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
 RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
 RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,  
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
 RT deltaH: functional analysis and comparative genomics.";  
 RL J. Bacteriol. 179:7135-7155(1997).

DR EMBL; AE000872; AAB85496.1; -.

DR InterPro; IPR000182; GCN5acetyltransf.

DR Pfam; PF00583; Acetyltransf; 1.

KW Transferase; Complete proteome.

SQ SEQUENCE 156 AA; 18162 MW; 910568F70B30A041 CRC64;

Query Match

Best Local Similarity 37.5%; Score 6; DB 17; Length 156;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIIFI 14

Db 53 FIIIFI 58  
 |||||

## RESULT 3

O25493 ID O25493 PRELIMINARY; PRT; 205 AA.  
 AC O25493;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein HP0813.  
 GN HP0813.

OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.

OX NCBI\_TaxID=210;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=26695 / ATCC 700392;

RX MEDLINE=97394467; PubMed=9252185;

RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Glodek A.,  
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Hickey E.K.,  
 RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,  
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 RA Venter J.C.;

RT "The complete genome sequence of the gastric pathogen Helicobacter  
 RT pylori.";  
 RL Nature 388:539-547(1997).

DR EMBL; AE000593; AAD07862.1; -.

DR TIGR; HP0813; -.

DR InterPro; IPR001279; Blactmase-like.

DR Pfam; PF00753; lactamase\_B; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 205 AA; 23463 MW; 9FD355E0B19C2DBE CRC64;

Query Match

Best Local Similarity 37.5%; Score 6; DB 16; Length 205;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11

|||||

Db 23 GIDFII 28

## RESULT 4

O8REQ8 ID O8REQ8 PRELIMINARY; PRT; 250 AA.  
 AC O8REQ8;

DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Branched-chain amino acid transport protein azlc.  
 GN FN1039.

OS Fusobacterium nucleatum (subsp. nucleatum).

OC Bacteria; Fusobacteria; Fusobacterium.

OX NCBI\_TaxID=76856;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 25586;

RX MEDLINE=21886394; PubMed=11889109;

RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
 RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,  
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
 RA Fonstein M., Kyrpides N., Overbeek R.;  
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
 RT nucleatum strain ATCC 25586.";  
 RL J. Bacteriol. 184:2005-2018(2002).



DR EMBL; AB010611; AAL95235.1; -.  
KW Complete proteome.  
SQ SEQUENCE 250 AA; 28752 MW; 42F8DF8D5F7432AF CRC64;

Query Match 37.5%; Score 6; DB 16; Length 250;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11  
| | | | |  
Db 179 GIDFII 184

## RESULT 5

Q9XYQ1 Q9XYQ1 PRELIMINARY; PRT; 319 AA.

AC Q9XYQ1;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Inositol 1,3,4-trisphosphate 5/6-kinase.  
GN IK.

OS Entamoeba histolytica.  
OC Eukaryota; Entamoebidae; Entamoeba.  
OX NCBI\_TaxID=5759;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=HM-1:IMSS;

RX MEDLINE=20264031; PubMed=10802324;  
RA Field J., Wilson M.P., Mai Z., Majerus P.W., Samuelson J.;

RT "An Entamoeba histolytica inositol 1,3,4-trisphosphate 5/6-kinase has  
a novel 3-kinase activity.";  
RL Mol. Biochem. Parasitol. 108:119-123(2000).  
DR EMBL; AF118848; AAD22969.1; -.

KW Kinase.  
SQ SEQUENCE 319 AA; 36480 MW; D526DBF2E897305D CRC64;

Query Match 37.5%; Score 6; DB 5; Length 319;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGIDFI 10  
| | | | |  
Db 272 CGIDFI 277

## RESULT 6

Q9N2T4 Q9N2T4 PRELIMINARY; PRT; 328 AA.

AC Q9N2T4;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Y9C9A.9 protein.  
GN Y9C9A.9.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;

RA None;

RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;

RA Waterston R.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC024882; AAF60933.1; -.

DR InterPro; IPR003002; 7TM\_chemol.  
DR InterPro; IPR000168; 7TM\_nematode.  
DR Pfam; PF01461; 7tm\_4; 1.  
SQ SEQUENCE 328 AA; 37879 MW; 86DD68EA42EB8FED CRC64;

Query Match 37.5%; Score 6; DB 5; Length 328;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWIF 15  
| | | | |  
Db 134 IIFWIF 139

## RESULT 7

Q23072 Q23072 PRELIMINARY; PRT; 330 AA.

ID Q23072;  
AC Q23072;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE ZC142.1 protein.  
GN ZC142.1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,  
Craxton M., Dear S., Du Z., Durbin R., Favello A., Jier M., Johnston L.,  
Gardner A., Green P., Hawkins T., Hillier L., Laister N., Latreille P.,  
Jones M., Kershaw J., Kirsten J., Murray A., Mortimore B., O'Callaghan M.,  
Lightning J., Lloyd C., McMuray A., Roopra A., Saunders D., Sulston J.,  
Parsons J., Percy C., Rifken L., Roopra A., Staden R., Sulston J.,  
Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;

RA Bradshaw H.;

RT "The sequence of C. elegans cosmid ZC142.";  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Waterston R.;

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U64841; AAB04845.1; -.

DR InterPro; IPR002651; DUF32.

DR Pfam; PF01748; DUF32; 1.

SQ SEQUENCE 330 AA; 37731 MW; 7C7399A117F7C463 CRC64;

## Query Match

Best Local Similarity 37.5%; Score 6; DB 5; Length 330;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWIF 15  
| | | | |  
Db 133 IIFWIF 138

## RESULT 8

O87850  
ID O87850 PRELIMINARY; PRT; 360 AA.  
AC O87850;

DT 01-NOV-1998 (TReMBLrel. 08, Created)  
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Putative secreted protein.  
GN SC06595 OR SC8A6.16.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL031013; CAA19788.1; --  
SQ SEQUENCE 360 AA; 38776 MW; BB1ED7F5BBAD8E7A CRC64;

Query Match 37.5%; Score 6; DB 16; Length 360;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANCG 6  
| | | | |  
Db 90 FOANCG 95

RESULT 9  
Q68409 PRELIMINARY; PRT; 399 AA.  
ID Q68409  
AC Q68409;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE Orf UL154.  
OS Human cytomegalovirus (strain Towne).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
OX NCBI\_TaxID=10363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TOWNE;  
RX MEDLINE=96099416; PubMed=8523595;  
RA Cha T.A., Tom E., Kemble G.W., Duke G.M., Mocarski E.S., Spaete R.R.;  
RT "Human cytomegalovirus clinical isolates carry at least 19 genes not  
RT found in laboratory strains."  
RL J. Virol. 70:78-83(1996).  
DR EMBL; U33332; AAA85896.1; --  
SQ SEQUENCE 399 AA; 45181 MW; 6D89F5267EF17998 CRC64;

Query Match 37.5%; Score 6; DB 12; Length 399;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFWI 14  
| | | | |  
Db 20 FIIFWI 25

RESULT 10  
Q9SMM7 PRELIMINARY; PRT; 430 AA.  
ID Q9SMM7  
AC Q9SMM7;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE Hypothetical 48.0 kDa protein.  
GN T8P19.230.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Choisme N., Robert C., Brottier P., Wincker P., Cattolico L.,  
RA Artiguenave F., Saurin W., Weissensbach J., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X., Quetier F., Salanoubat M.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBSJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; AL133315; CAB62361.1; --  
DR InterPro; IPR003480; Transferase.  
DR Pfam; PF02458; Transferase; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 430 AA; 48004 MW; 6210941B7C148B31 CRC64;

Query Match 37.5%; Score 6; DB 10; Length 430;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ANCGID 8  
| | | | |  
Db 103 ANCGID 108

RESULT 11  
Q9WI36 PRELIMINARY; PRT; 674 AA.  
ID Q9WI36  
AC Q9WI36;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE Reverse transcriptase.  
OS Cauliflower mosaic virus.  
OC Viruses; Retroid viruses; Caulimoviridae; Caulimovirus.  
OX NCBI\_TaxID=10641;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=XINJIANG;  
RA Fang R., Wu X., Bu M., Tian Y., Cai F., Mang K.;  
RT "Complete nucleotide sequence of cauliflower mosaic virus (Xinjiang  
RT isolate) genomic DNA."  
RL Ping Tu Hsueh Pao 1:247-256(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=XINJIANG;  
RA Fang R., Wu X., Bu M., Tian Y., Cai F., Mang K.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; AF140604; AAD37341.1; --  
DR MEROPS; A03.001; --  
DR InterPro; IPR000588; Peptidase\_A3.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF02160; Peptidase\_A3; 1.  
DR Pfam; PF00078; rvt; 1.  
DR PRINTS; PR00731; CAULIMOPTASE.  
KW RNA-directed DNA polymerase.  
SQ SEQUENCE 674 AA; 78176 MW; 7C9AE48ACDA4D205 CRC64;

Query Match 37.5%; Score 6; DB 12; Length 674;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11  
| | | | |

Db 111 GIDFII 116

## RESULT 12

Q83169  
ID Q83169 PRELIMINARY; PRT; 680 AA.  
AC Q83169;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Reverse transcriptase.  
OS Cauliflower mosaic virus.  
OC Viruses; Retroid viruses; Caulimoviridae; Caulimovirus.  
OX NCBI\_TaxID=10641;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94143488; PubMed=8310068;  
RA Chenault K.D., Melcher U.;  
RT "Cauliflower mosaic virus isolate CMV-1.";  
RL Plant Physiol. 101:1395-1396(1993).  
DR EMBL; M90543; AAA21736.1; -.  
DR MEROPS; A03.001; -.  
DR InterPro; IPR000588; Peptidase\_A3.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF02160; Peptidase\_A3; 1.  
DR Pfam; PF00078; rvt; 1.  
DR PRINTS; PR00731; CAULIMOPTASE.  
DR RNA-directed DNA polymerase.  
KW RNA-directed DNA polymerase.  
SQ SEQUENCE 680 AA; 78727 MW; ACA8A4B9E8316708 CRC64;

Query Match 37.5%; Score 6; DB 12; Length 680;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11

|||||

Db 110 GIDFII 115

## RESULT 13

Q66162  
ID Q66162 PRELIMINARY; PRT; 680 AA.  
AC Q66162;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ORF V.  
OS Cauliflower mosaic virus.  
OC Viruses; Retroid viruses; Caulimoviridae; Caulimovirus.  
OX NCBI\_TaxID=10641;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B29;  
RX MEDLINE=95237629; PubMed=7721109;  
RA Pique M., Mougeot J.L., Geldreich A., Guidasci T., Mesnard J.M.,  
RA Lebeurier G., Yot P.;  
RT "Sequence of a cauliflower mosaic virus strain infecting solanaceous  
RT plants.";  
RL Gene 155:305-306(1995).  
DR EMBL; X79465; CAA55974.1; -.  
DR MEROPS; A03.001; -.  
DR InterPro; IPR000588; Peptidase\_A3.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF02160; Peptidase\_A3; 1.  
DR Pfam; PF00078; rvt; 1.  
DR PRINTS; PR00731; CAULIMOPTASE.  
DR RNA-directed DNA polymerase.  
KW RNA-directed DNA polymerase.  
SQ SEQUENCE 680 AA; 78728 MW; CF1EDBBEF45EC2A6 CRC64;

Query Match 37.5%; Score 6; DB 12; Length 680;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11

|||||

Db 110 GIDFII 115

## RESULT 14

Q9KAQ7  
ID Q9KAQ7 PRELIMINARY; PRT; 1091 AA.  
AC Q9KAQ7;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Hypothetical protein BH2230.  
GN BH2230.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL; AP001514; BAB05949.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1091 AA; 127442 MW; FD95071E13701331 CRC64;

Query Match 37.5%; Score 6; DB 16; Length 1091;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11

|||||

Db 880 GIDFII 885

## RESULT 15

Q8V3M7  
ID Q8V3M7 PRELIMINARY; PRT; 1285 AA.  
AC Q8V3M7;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE SPV068 RNA polymerase subunit RPO147.  
GN SPV068.  
OS Swinepox virus.  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Suipoxvirus.  
OX NCBI\_TaxID=10276;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=17077-99;  
RX MEDLINE=21624277; PubMed=11752168;  
RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Osorio F.A., Balinsky C.,  
RA Kutish G.F., Rock D.L.;  
RT "The genome of swinepox virus.";  
RL J. Virol. 76:783-790(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=17077-99;  
RA Afonso C.L., Tulman E.R., Lu Z., Balinsky C., Osorio F.A., Zsak L.,  
RA Kutish G.F., Rock D.L.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF410153; AAL69807.1; -.  
DR InterPro; IPR000722; RNA\_pol\_A.  
DR Pfam; PF00623; RNA\_pol\_A; 1.  
SQ SEQUENCE 1285 AA; 147680 MW; 1D34EFF8CD873DC6 CRC64;

Query Match 37.5%; Score 6; DB 12; Length 1285;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11  
|11111|  
Db 552 GIDFII 557

## RESULT 16

Q9PIF6 Q9PIF6 PRELIMINARY; PRT; 41 AA.  
AC Q9PIF6;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE Hypothetical protein Cj0344.  
GN Cj0344.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
OC Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCTC 11168;  
RX MEDLINE=20150912; PubMed=10688204;  
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
BAsham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,  
Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
Whitehead S., Barrall B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
reveals hypervariable sequences.";  
RT Nature 403:665-668(2000).  
RL EMBL; AL139075; CAB74181.1; -  
DR Hypothetical protein; Complete proteome.  
KW SEQUENCE 41 AA; 5107 MW; 2698D105B0F0BF04 CRC64;  
SQ

Query Match 31.2%; Score 5; DB 16; Length 41;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12  
|11111|  
Db 10 DFIIIF 14

## RESULT 17

Q9ZZU6 Q9ZZU6 PRELIMINARY; PRT; 52 AA.  
AC Q9ZZU6;  
DT 01-MAY-1999 (TReMBLrel. 10, Created)  
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE ATPase subunit 8.  
GN ATPASE 8.  
OS Littorina saxatilis.  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;  
OC Mesogastropoda; Littorinoidea; Littorinidae; Littorina.  
OX NCBI\_TaxID=31220;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99193185; PubMed=10093225;  
RA Wilding C.S., Mill P.J., Grahame J.;  
RT "Partial sequence of the mitochondrial genome of Littorina saxatilis:  
relevance to gastropod phylogenetics.";  
RL J. Mol. Evol. 48:348-359(1999).  
DR EMBL; AJ132137; CAAL0595.1; -  
DR InterPro; IPR001421; ATPase8\_mit.  
DR Pfam; PF00895; ATP-synt\_8; 1.  
KW Mitochondrion.  
SQ SEQUENCE 52 AA; 6237 MW; 756DDE254F629499 CRC64;

Query Match 31.2%; Score 5; DB 8; Length 52;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIIFW 13  
|11111|  
Db 13 FIIIFW 17

## RESULT 18

Q8RH01 Q8RH01 PRELIMINARY; PRT; 57 AA.  
AC Q8RH01;  
DT 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Hypothetical protein FN0120.  
GN FN0120.  
OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteria; Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25586;  
RX MEDLINE=21886394; PubMed=11889109;  
RA Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
BHattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,  
Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
Fonstein M., Kyrpides N., Overbeek R.;  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
nucleatum strain ATCC 25586.";  
RL J. Bacteriol. 184:2005-2018(2002).  
DR EMBL; AE010526; AAL94329.1; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 57 AA; 6631 MW; EF675C5C73A500BE CRC64;

Query Match 31.2%; Score 5; DB 16; Length 57;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|11111|  
Db 21 GIDFI 25

## RESULT 19

Q9AIF7 Q9AIF7 PRELIMINARY; PRT; 83 AA.  
AC Q9AIF7;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Ribosomal protein S17.  
GN RPS17.  
OS Carsonella ruddii.  
OC Bacteria; Proteobacteria; gamma subdivision; Candidatus Carsonella.  
OX NCBI\_TaxID=114186;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21125546; PubMed=11222582;  
RA Clark M.A., Baumann L., Thao M.L., Moran N.A., Baumann P.;  
RT "Degenerative Minimalism in the Genome of a Psyllid Endosymbiont.";  
RL J. Bacteriol. 183:1853-1861(2001).  
DR EMBL; AF274444; AAK17089.1; -  
DR HSSP; P23828; IRIP.  
DR InterPro; IPR000266; Ribosomal\_S17.  
DR Pfam; PF00366; Ribosomal\_S17; 1.  
DR ProDom; PD001295; Ribosomal\_S17; 1.  
SQ SEQUENCE 83 AA; 10205 MW; 72EA2AC8F1AC09A8 CRC64;

Query Match 31.2%; Score 5; DB 2; Length 83;



Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0; Matches 5; Conservative 0; Mismatches 0;

QY 10 IIFWI 14  
|  
|  
|  
|  
Db 68 IIFWI 72

RESULT 20  
Q8TMA6 PRELIMINARY; PRT; 90 AA.  
AC Q8TMA6;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Predicted protein.  
GN MA2759.

OS Methanosarcina acetivorans.  
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
OC Methanosarcinaceae; Methanosarcina.  
OX NCBI\_TaxID=2214;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
RX MEDLINE=21929760; PubMed=11932238;  
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,  
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,  
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,  
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,  
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,  
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,  
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
RA Metcalf W.W., Birren B.;  
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic  
RT and physiological diversity.";  
RL Genome Res. 12:532-542(2002).  
DR EMBL; AE010975; AAM06137.1; -.  
KW Complete proteome.  
SQ SEQUENCE 90 AA; 10457 MW; DDD0D4BC3F000CD8 CRC64;

Query Match 31.2%; Score 5; DB 17; Length 90;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0; Matches 5; Conservative 0; Mismatches 0;

QY 9 FIIFW 13  
|  
|  
|  
|  
Db 75 FIIFW 79

RESULT 21  
Q9Z5P6 PRELIMINARY; PRT; 103 AA.  
ID Q9Z5P6  
AC Q9Z5P6;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE Hypothetical 12.0 kDa protein (Traf protein).  
GN EP0045.

OS Enterococcus faecalis (Streptococcus faecalis).  
OG Plasmid pAM373.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Enterococcaceae; Enterococcus.  
OX NCBI\_TaxID=1351;  
RN [1]  
RP SEQUENCE FROM N.A.

RC PLASMID=PAM373;  
RA Muscholl-Silberhorn A.B.;  
RT "Cloning and expression of Asa373, a novel adhesin unrelated to the  
RT other sex pheromone plasmid-encoded aggregation substances of  
RT Enterococcus faecalis.";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC PLASMID=PAM373;  
RX MEDLINE=20453452; PubMed=10998166;  
RA De Boever E.H., Clewell D.B., Fraser C.M.;  
RT "Enterococcus faecalis conjugative plasmid pAM373: complete nucleotide  
RT sequence and genetic analyses of sex pheromone response.";  
RL Mol. Microbiol. 37:1327-1341(2000).  
DR EMBL; AJ132039; CAB38224.1; -.  
DR EMBL; AE002565; AAG40456.1; -.  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 103 AA; 12019 MW; 91F4F000B80B2D35 CRC64;

Query Match 31.2%; Score 5; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0; Matches 5; Conservative 0; Mismatches 0;

QY 4 NCGID 8  
|  
|  
|  
|  
Db 26 NCGID 30

RESULT 22  
Q55541 PRELIMINARY; PRT; 106 AA.  
ID Q55541  
AC Q55541;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein sir0333.  
GN SLR0333.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PCC 6803;  
RA Tabata S.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.

RX MEDLINE=96127529; PubMed=8590279;  
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
RA Sugiyura M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
RT region from map positions 64% to 92% of the genome.";  
RL DNA Res. 2:153-166(1995).  
RN [3]  
RP SEQUENCE FROM N.A.

RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hirose M., Sugiyura M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
RA Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
DR EMBL; D63999; BAA10038.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 106 AA; 11406 MW; EEE4CF44B5B79C3B CRC64;

Query Match 31.2%; Score 5; DB 16; Length 106;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0; Matches 5; Conservative 0; Mismatches 0;

QY 8 DFIF 12  
|  
|  
|  
|  
Db 97 DFIF 101

## RESULT 23

Q9S6Y4  
ID Q9S6Y4 PRELIMINARY; PRT; 107 AA.  
AC Q9S6Y4;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Dihydroflavanol reductase 3 (Fragment).  
GN DFR.  
OS Lotus corniculatus (Bird's-foot trefoil).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.  
OX NCBI\_TaxID=47247;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LEAF;  
RA Paolocci F., Capucci R., Arcioni S., Damiani F.;  
RT "Birdsfoot trefoil: a model for studying the synthesis of condensed  
RT tannin.";  
RL (In) Gross G.G., Hemingway R.W., Yoshida T. (eds.);  
RL PLANT POLYPHENOLS 2, CHEMISTRY AND BIOLOGY, pp.1-1, Plenum Press,  
RL New York, NY, USA (2000).  
DR EMBL; AF117263; AAF23884.1; -.  
FT NON\_TER 1 107  
FT NON\_TER 107 107  
SQ SEQUENCE 107 AA; 12242 MW; C5BC292A44246D7D CRC64;

Query Match 31.2%; Score 5; DB 10; Length 107;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 86 GIDFI 90

## RESULT 24

Q9S6Y3  
ID Q9S6Y3 PRELIMINARY; PRT; 107 AA.  
AC Q9S6Y3;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Dihydroflavanol reductase 4 (Fragment).  
GN DFR.  
OS Lotus corniculatus (Bird's-foot trefoil).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.  
OX NCBI\_TaxID=47247;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LEAF;  
RA Paolocci F., Capucci R., Arcioni S., Damiani F.;  
RT "Birdsfoot trefoil: a model for studying the synthesis of condensed  
RT tannin.";  
RL (In) Gross G.G., Hemingway R.W., Yoshida T. (eds.);  
RL PLANT POLYPHENOLS 2, CHEMISTRY AND BIOLOGY, pp.1-1, Plenum Press,  
RL New York, NY, USA (2000).  
DR EMBL; AF117264; AAF23885.1; -.  
FT NON\_TER 1 107  
FT NON\_TER 107 107  
SQ SEQUENCE 107 AA; 12170 MW; C6EALC4A44246D7D CRC64;

Query Match 31.2%; Score 5; DB 10; Length 107;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 86 GIDFI 90

## RESULT 25

Q936F3  
ID Q936F3 PRELIMINARY; PRT; 108 AA.  
AC Q936F3;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Hypothetical 12.2 kDa protein.  
OS Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Staphylococcus.  
OX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M;  
RA Luong T.T., Shu O., Bush K., Lee C.Y.;  
RT "The Tygel Capsular Polysaccharide of Staphylococcus aureus is carried  
RT in a Staphylococcal Cassette Chromosome Genetic Element.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U10927; AAL26685.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 108 AA; 12234 MW; 40380A7E555A40DB CRC64;

Query Match 31.2%; Score 5; DB 2; Length 108;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 34 GIDFI 38

## RESULT 26

Q8VVE1  
ID Q8VVE1 PRELIMINARY; PRT; 113 AA.  
AC Q8VVE1;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Ssb protein.  
GN SSB.  
OS uncultured bacterium.  
OG Plasmid pEMT3.  
OC Bacteria; environmental samples.  
OX NCBI\_TaxID=77133;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gstalder M.E.;  
RT "Caracterisation de plasmides a large spectre d'notes isoles de  
RL biotopes pollues.";  
RL Thesis (2001), Department of Biological sciences,  
RL Universite Libre de Bruxelles, Bruxelles, Belgium.  
DR EMBL; AJ414161; CAC94917.1; -.  
DR InterPro; IPR000424; SSB\_protein.  
DR Pfam; PF00436; SSB; 1.  
DR TIGRFAMs; TIGR00621; ssb; 1.  
KW Plasmid.  
SQ SEQUENCE 113 AA; 12836 MW; 9F5D88B9ABF476FE CRC64;

Query Match 31.2%; Score 5; DB 2; Length 113;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 94 GIDFI 98

## RESULT 27

Q9FDG9  
ID Q9FDG9 PRELIMINARY; PRT; 118 AA.

```
AC Q9FDG9;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Hypothetical 13.5 kDa protein.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH43;
RX MEDLINE=20378620; PubMed=10919773;
RA Qi F., Chen P., Caufield P.W.;
RT "Purification and biochemical characterization of mutacin I from the
RT group I strain of streptococcus mutans, CH43, and genetic analysis of
RT mutacin I biosynthesis genes.";
RL Appl. Environ. Microbiol. 66:3221-3229(2000).
DR EMBL; AF267498; AAF99696.1; -.
KW Hypothetical protein.
SQ SEQUENCE 118 AA; 13465 MW; C1D9590A77CF4D4D CRC64;

Query Match 31.2%; Score 5; DB 2; Length 118;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 IFWIF 15
Db 14 IFWIF 18

RESULT 28
Q8Y6N6 PRELIMINARY; PRT; 120 AA.
AC Q8Y6N6;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein lmol648.
GN LM01648.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madsen E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591980; CAC99726.1; -.
DR ListiList; LM001648; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 120 AA; 14246 MW; AE139D9B67330B8F CRC64;

Query Match 31.2%; Score 5; DB 16; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11
Db 27 IDFI 31
```

```
RESULT 29
O68605 PRELIMINARY; PRT; 123 AA.
ID O68605;
AC O68605;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 13.5 kDa protein (Fragment).
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC11168;
RX MEDLINE=98440418; PubMed=9765558;
RA van Vliet A.H.M., Wooldridge K.G., Ketley J.M.;
RT "Iron-responsive gene regulation in a Campylobacter jejuni fur
RT mutant.";
RL J. Bacteriol. 180:5291-5298(1998).
DR EMBL; AF052056; AAC64261.1; -.
DR InterPro; IPR003825; Colicin_V.
DR Pfam; PF02674; Colicin_V; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 123 AA; 13508 MW; 2B2F9B250E4B63A7 CRC64;

Query Match 31.2%; Score 5; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWI 14
Db 13 IIFWI 17

RESULT 30
O95545 PRELIMINARY; PRT; 126 AA.
ID O95545;
AC O95545;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE DJ283E3.6.3 (Putative novel protein similar to many (Archae)bacterial,
DE worm and yeast hypothetical proteins) (Putative partial isoform 3)
DE (Fragment).
DE DJ283E3.6.
GN DJ283E3.6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pearce A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031282; CAA20355.1; -.
DR InterPro; IPR002504; ATP_NADK.
DR Pfam; PF01513; NAD_kinase; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 126 AA; 14188 MW; 79D0000A1E5B499B1 CRC64;

Query Match 31.2%; Score 5; DB 4; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11
Db 96 IDFI 100
```

RESULT 31  
Q53372  
ID Q53372 PRELIMINARY; PRT; 128 AA.  
AC Q53372;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE C-type cytochrome.  
GN CYTM.  
OS Synechocystis sp. (strain PCC 6714).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1147;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PCC6714;  
RA Malakhov M.P., Semenenko V.E.;  
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X82563; CAA57910.1; -.  
DR HSP; P00044; ICRI.  
DR InterPro; IPR000345; CytC\_heme\_bind.  
DR InterPro; IPR003088; Cyt\_C1.  
DR Pfam; PF00034; cytochrome\_c; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
FT CHAIN 53 128 POTENTIAL.  
SQ SEQUENCE 128 AA; 13848 MW; D7725CD852C25C01 CRC64;

Query Match  
Best Local Similarity 31.2%; Score 5; DB 2; Length 128;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FQANC 5  
Db 64 FQANC 68

RESULT 32  
O26942  
ID O26942 PRELIMINARY; PRT; 128 AA.  
AC O26942;  
DT 01-JAN-1998 (TReMBLrel. 05, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE FERREDOXIN.  
GN MTH854.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DELTA H;  
RX MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
deltaH: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155(1997).  
DR EMBL; AE000862; AAB85352.1; -.  
DR HSP; P00198; 2FDN.  
DR InterPro; IPR001450; 4Fe4S\_ferredoxin.  
DR Pfam; PF00037; fer4; 2.  
DR PRINTS; PR00353; 4FE4SFRDOXIN.  
DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 2.  
KW Iron-sulfur; Complete proteome.  
SQ SEQUENCE 128 AA; 14216 MW; 4EEB2C4261D5BE9F CRC64;

Query Match  
Best Local Similarity 31.2%; Score 5; DB 17; Length 128;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 GIDFI 10  
Db 54 GIDFI 58  
RESULT 33  
Q9HK39  
ID Q9HK39 PRELIMINARY; PRT; 134 AA.  
AC Q9HK39;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE Hypothetical protein Ta0768.  
GN TA0768.  
OS Thermoplasma acidophilum.  
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmales;  
OC Thermoplasmataceae; Thermoplasma.  
OX NCBI\_TaxID=2303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 1728;  
RX MEDLINE=20479972; PubMed=11029001;  
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;  
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma  
acidophilum.";  
RL Nature 407:508-513(2000).  
DR EMBL; AL445065; CAC11900.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 134 AA; 15217 MW; 5A3A02E8FB8FF4FE CRC64;

Query Match  
Best Local Similarity 31.2%; Score 5; DB 17; Length 134;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 127 FIIFW 131

RESULT 34  
Q13964  
ID Q13964 PRELIMINARY; PRT; 136 AA.  
AC Q13964;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)  
DE Glycoprotein CD28.  
GN CD28.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90293482; PubMed=2162892;  
RA Lee K.P., Taylor C., Petryniak B., Turka L.A., June C.H.,  
RA Thompson C.B.;  
RT "The genomic organization of the CD28 gene. Implications for the  
regulation of CD28 mRNA expression and heterogeneity.";  
RL J. Immunol. 145:344-352(1990).  
DR EMBL; M37815; AAA51945.1; -.  
DR EMBL; M37812; AAA51945.1; JOINED.  
DR EMBL; M37813; AAA51945.1; JOINED.  
DR EMBL; M37814; AAA51945.1; JOINED.  
SQ SEQUENCE 136 AA; 15369 MW; C9AF33467706D2BE CRC64;

Query Match  
Best Local Similarity 31.2%; Score 5; DB 4; Length 136;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 9 FIFW 13  
Db 90 FIFW 94  
|||||  
RESULT 35  
Q21953 PRELIMINARY; PRT; 136 AA.  
AC Q21953;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Hypothetical 15.2 kDa protein.  
GN R12B2.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium."  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Miller N.;  
RT "The sequence of C. elegans cosmid R12B2.";  
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission."  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U00066; AAA50741.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 136 AA; 15226 MW; 611D037F55CA7696 CRC64;  
Query Match 31.2%; Score 5; DB 5; Length 136;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 DFIF 12  
Db 25 DFIF 29  
|||||  
RESULT 36  
Q97KA8 PRELIMINARY; PRT; 138 AA.  
AC Q97KA8;  
DT 01-OCT-2001 (TREMBlrel. 18, Created)  
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
DE Uncharacterized small membrane protein, homolog of ykva  
B.subtilis.  
DE B.subtilis.  
GN CAC1011.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1488;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=21359325; PubMed=11466286;  
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
RA Bennett G.N., Koonin E.V., Smith D.R.;

RT "Genome sequence and comparative analysis of the solvent-producing  
bacterium Clostridium acetobutylicum.";  
RL J. Bacteriol. 183:4823-4838(2001).  
DR EMBL; AE007616; AAK78987.1; -.  
KW Complete proteome.  
SQ SEQUENCE 138 AA; 16257 MW; 570837AF9060ACC5 CRC64;  
Query Match 31.2%; Score 5; DB 16; Length 138;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 10 IIFWI 14  
Db 117 IIFWI 121  
|||||  
RESULT 37  
Q974Q5 PRELIMINARY; PRT; 147 AA.  
ID Q974Q5  
AC Q974Q5;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical protein ST0604.  
GN ST0604.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=111955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JCM 10545 / 7;  
RX PubMed=11572479;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,  
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,  
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.;  
RT "Complete genome sequence of an aerobic thermoacidophilic  
Crenarchaeon, Sulfolobus tokodaii strain7.";  
RL DNA Res. 8:123-140(2001).  
DR EMBL; AP000983; BAB65602.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 147 AA; 17362 MW; DE2A7D819E08A3C6 CRC64;  
Query Match 31.2%; Score 5; DB 17; Length 147;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 DFIF 12  
Db 36 DFIF 40  
|||||  
RESULT 38  
Q9NWH4 PRELIMINARY; PRT; 148 AA.  
ID Q9NWH4  
AC Q9NWH4;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Hypothetical 16.9 kDa protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EMBRYO;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.;

RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;  
RT "NEDO human cDNA sequencing project."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK000886; BAA91406.1; -  
DR InterPro; IPR001081; Ribosomal\_L20.  
DR ProDom; PD002389; Ribosomal\_L20; 1.  
SQ SEQUENCE 148 AA; 16903 MW; 9A28D55AEDED4766 CRC64;

Query Match 31.2%; Score 5; DB 4; Length 148;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QANCG 6  
|||||  
Db 117 QANCG 121

RESULT 39  
Q8REG9  
ID Q8REG9 PRELIMINARY; PRT; 155 AA.  
AC Q8REG9;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical cytosolic protein FN1134.  
GN FN1134.

OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteria; Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25586;  
RX MEDLINE=21886394; PubMed=11889109;  
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
RA Bhatnagar V., Bartman A., Gardner W., Grechkin G., Zhu L.,  
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,  
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
RA Fonstein M., Kyrpides N., Overbeek R.;  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
RT nucleatum strain ATCC 25586."  
RL J. Bacteriol. 184:2005-2018(2002).  
DR EMBL; AE010618; AAL95330.1; -  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 155 AA; 18101 MW; CB3EDB741E897B97 CRC64;

Query Match 31.2%; Score 5; DB 16; Length 155;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11  
|||||  
Db 27 IDFI 31

RESULT 40  
Q9RFAL  
ID Q9RFAL PRELIMINARY; PRT; 158 AA.  
AC Q9RFAL;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical 17.3 kDa protein (Fragment).  
OS Salmonella pullorum.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=605;  
RN [1]

RP SEQUENCE FROM N.A.  
RA Patton T., Tseng C.W., Sobieski R.J., Crupper S.S.;  
RT "Salmonella pullorum chromosomal region demonstrating homology to  
RT Escherichia coli K-12 MG1655 section 398 to 400 of the genome."  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF198630; AAF15288.1; -  
DR InterPro; IPR002599; Metalloenzyme.  
DR Pfam; PF01676; Metalloenzyme; 1.  
KW Hypothetical protein.  
FT NON\_TER 1  
FT NON\_TER 158  
SQ SEQUENCE 158 AA; 17266 MW; 96F489ADBDFEB862 CRC64;

Query Match 31.2%; Score 5; DB 2; Length 158;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ANCGI 7  
|||||  
Db 31 ANCGI 35

RESULT 41  
Q97L62  
ID Q97L62 PRELIMINARY; PRT; 160 AA.  
AC Q97L62;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Predicted tRNA-methylase (SpoV class).  
GN CAC0700.

OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1488;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=21359325; PubMed=11466286;  
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
RA Bennett G.N., Koonin E.V., Smith D.R.;  
RT "Genome sequence and comparative analysis of the solvent-producing  
RT bacterium Clostridium acetobutylicum."  
RL J. Bacteriol. 183:4823-4838(2001).  
DR EMBL; AE007585; AAK78677.1; -  
DR InterPro; IPR001537; SpoU\_methylase.  
DR Pfam; PF00588; SpoU\_methylase; 1.  
KW Methyltransferase; Complete proteome.

SQ SEQUENCE 160 AA; 18424 MW; 2F6812B632847820 CRC64;

Query Match 31.2%; Score 5; DB 16; Length 160;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12  
|||||  
Db 101 DFIIIF 105

RESULT 42  
Q8XKN7  
ID Q8XKN7 PRELIMINARY; PRT; 161 AA.  
AC Q8XKN7;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Probable transcriptional regulator.  
GN CPE1357.

OS Clostridium perfringens.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1502;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=13 / TYPE A;  
RX PubMed=11792842;

RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
RT flesh-eater."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
DR EMBL; AP003190; BAB81063.1; -.  
KW Complete proteome.  
SQ SEQUENCE 161 AA; 17955 MW; 23772A02B24563B2 CRC64;

Query Match 31.2%; Score 5; DB 16; Length 161;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
Db 128 DFIIF 132

RESULT 43  
Q8YPA6 PRELIMINARY; PRT; 163 AA.  
AC Q8YPA6;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein Alr4293.  
GN ALR4293.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Irliguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shlimpo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120."  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003596; BAB75992.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 163 AA; 18352 MW; AF0D6B94CACD6806 CRC64;

Query Match 31.2%; Score 5; DB 16; Length 163;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 IFWIF 15  
Db 58 IFWIF 62

RESULT 44  
Q99UW3 PRELIMINARY; PRT; 173 AA.  
AC Q99UW3;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein SAV1142.  
GN SAV1142 OR SA0989.  
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
OS Staphylococcus aureus (strain N315).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Staphylococcus.  
OX NCBI\_TaxID=158878, 158879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,

RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
RT aureus."  
RL Lancet 357:1225-1240(2001).  
DR EMBL; AP003361; BAB57304.1; -.  
DR EMBL; AP003132; BAB42238.1; -.  
DR InterPro; IPR003825; Colicin\_V.  
DR Pfam; PF02674; Colicin\_V; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 173 AA; 20286 MW; E8AD9D5C728EB44E CRC64;

Query Match 31.2%; Score 5; DB 16; Length 173;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
Db 3 IDFII 7

RESULT 45  
Q8WXJ2 PRELIMINARY; PRT; 178 AA.  
ID Q8WXJ2  
AC Q8WXJ2;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE CD28 antigen (Fragment).  
GN CD28.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21601151; PubMed=11735222;  
RA Ling V., Wu P.W., Finnerty H.F., Agostino M.J., Graham J.R., Chen S.,  
RA Jussiff J., Fisk G.J., Miller C.P., Collins M.;  
RT "Assembly and Annotation of Human Chromosome 2q33 Sequence Containing  
RT the CD28, CTLA4, and ICOS Gene Cluster: Analysis by Computational,  
RT Comparative, and Microarray Approaches."  
RL Genomics 78:155-168(2001).  
DR EMBL; AF411057; AAL40931.1; -.  
DR InterPro; IPR003600; Ig\_like.  
DR SMART; SM00410; Ig\_like; 1.  
FT NON\_TER 178  
SQ SEQUENCE 178 AA; 20022 MW; D60EBA4174AAABD0 CRC64;

Query Match 31.2%; Score 5; DB 4; Length 178;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 174 FIIFW 178

Search completed: May 11, 2003, 20:12:12  
Job time : 11.4956 secs





GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 13:47:50 ; Search time 362.761 Seconds  
(without alignments)  
1283.613 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 98  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues 4109280  
Total number of hits satisfying chosen parameters:  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_173955\_334/app\_query.fasta\_1.462  
-DB=GenEmbl -Qfmt=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn\_1\_1758\_@runat\_05052003\_173955\_334 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	98	100.0	337	6	AX321911	Sequence
2	98	100.0	2239	6	AX321909	Sequence
3	98	100.0	5981	6	AX321910	Sequence
c 4	98	100.0	161280	9	AC093903	Homo sapi
5	66	67.3	41179	3	AF003739	Caenorhab
6	66	67.3	201746	2	AC006895	Caenorhab
7	61	62.2	24259	6	AX251448	Sequence
8	60	61.2	94703	2	AC008563	Homo sapi
9	60	61.2	110879	9	AC016603	Homo sapi
c 10	60	61.2	204843	9	AC010359	Homo sapi
c 11	59	60.2	67679	2	AC113545	Mus muscu
12	59	60.2	161334	10	AC091629	Mus muscu
13	59	60.2	198281	10	AL593843	Mouse DNA
c 14	59	60.2	218188	2	AL845270	Mus muscu
c 15	57	58.2	149765	9	AC004984	Homo sapi
16	57	58.2	177147	9	AC007539	Homo sapi
17	57	58.2	191844	2	AC122410	Mus muscu
c 18	57	58.2	234096	2	AC121818	Mus muscu
19	56.5	57.7	236925	2	AC102650	Mus muscu
20	56	57.1	1669	10	BC009142	Mus muscu
c 21	56	57.1	22211	9	AL359849	Human DNA
22	56	57.1	47887	2	AC015109	Drosophil
23	56	57.1	52356	2	AC020166	Drosophil
c 24	56	57.1	119633	2	AC119679	Homo sapi
c 25	56	57.1	139394	2	AC010046	Drosophil
c 26	56	57.1	158098	10	AL603664	Mouse DNA
c 27	56	57.1	165330	2	AC016011	Homo sapi
28	56	57.1	171376	3	AC007929	Drosophil
c 29	56	57.1	174673	9	AC019187	Homo sapi
c 30	56	57.1	179055	3	AC093551	Drosophil
c 31	56	57.1	187192	2	AC128532	Rattus no
32	56	57.1	187756	2	AC108826	Mus muscu
33	56	57.1	190153	3	AC008201	Drosophil
34	56	57.1	191240	2	AC027738	Homo sapi
35	56	57.1	199851	2	AC125080	Mus muscu
36	56	57.1	203926	2	AC011122	Homo sapi
c 37	56	57.1	204207	2	AC023147	Homo sapi
38	56	57.1	220842	3	AE003745	Drosophil
39	56	57.1	232719	2	AC111123	Mus muscu
c 40	56	57.1	322667	3	AE003592	Drosophil
41	55.5	56.6	6526	6	AX344899	Sequence
42	55.5	56.6	6526	6	AX347139	Sequence
43	55.5	56.6	156146	2	AL139180	Homo sapi
44	55.5	56.6	180680	9	AL137141	Human DNA
c 45	55.5	56.6	211624	10	AC114817	Mus muscu

ALIGNMENTS

AX321911  
LOCUS AX321911 337 bp DNA linear PAT 15-DEC-2001  
DEFINITION Sequence 442 from Patent WO0172295.  
ACCESSION AX321911  
VERSION AX321911.1 GI:17906521  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Reed, S.G., Lodes, M.J., Mohamath, R., Secrist, H., Benson, D.R.,  
Indirias, C.Y., Henderson, R.A., Fling, S.P., Algate, P.A., Elliot, M.,  
Mannion, J. and Kalos, M.D.  
TITLE Compositions and methods for the therapy and diagnosis of lung  
cancer  
JOURNAL Patent: WO 0172295-A 442 04-OCT-2001;  
CORIXA CORPORATION (US)  
FEATURES  
source Location/Qualifiers  
1 .337  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 103 a 60 c 93 g 81 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.88e-08 Length: 337  
Score: 98.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
US-09-854-133-587 (1-16) x AX321911 (1-337)  
QY 1 PheGlnAlaAsnCysGlyIleAspPheIlePheTrpPheTrp 16  
|||||  
Db 107 TTCCAGGCCAATTGTGGCATAGATTTATCATATTTCTGGATTTTGG 154  
|||||  
RESULT 2  
AX321909  
LOCUS AX321909 2239 bp DNA linear PAT 15-DEC-2001  
DEFINITION Sequence 440 from Patent WO0172295.  
ACCESSION AX321909  
VERSION AX321909.1 GI:17906515  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Reed, S.G., Lodes, M.J., Mohamath, R., Secrist, H., Benson, D.R.,  
Indirias, C.Y., Henderson, R.A., Fling, S.P., Algate, P.A., Elliot, M.,  
Mannion, J. and Kalos, M.D.  
TITLE Compositions and methods for the therapy and diagnosis of lung  
cancer  
JOURNAL Patent: WO 0172295-A 440 04-OCT-2001;  
CORIXA CORPORATION (US)  
FEATURES  
source Location/Qualifiers  
1 .2239  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 619 a 444 c 493 g 683 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.5e-07 Length: 2239  
Score: 98.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX321909 (1-2239)  
QY 1 PheGlnAlaAsnCysGlyIleAspPheIlePheTrpPheTrp 16  
|||||  
Db 104 TTCCAGGCCAATTGTGGCATAGATTTATCATATTTCTGGATTTTGG 151  
|||||  
RESULT 3  
AX321910  
LOCUS AX321910 5981 bp DNA linear PAT 15-DEC-2001  
DEFINITION Sequence 441 from Patent WO0172295.  
ACCESSION AX321910  
VERSION AX321910.1 GI:17906518  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Reed, S.G., Lodes, M.J., Mohamath, R., Secrist, H., Benson, D.R.,  
Indirias, C.Y., Henderson, R.A., Fling, S.P., Algate, P.A., Elliot, M.,  
Mannion, J. and Kalos, M.D.  
TITLE Compositions and methods for the therapy and diagnosis of lung  
cancer  
JOURNAL Patent: WO 0172295-A 441 04-OCT-2001;  
CORIXA CORPORATION (US)  
FEATURES  
source Location/Qualifiers  
1 .5981  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 1858 a 1029 c 1098 g 1996 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7.68e-07 Length: 5981  
Score: 98.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
US-09-854-133-587 (1-16) x AX321910 (1-5981)  
QY 1 PheGlnAlaAsnCysGlyIleAspPheIlePheTrpPheTrp 16  
|||||  
Db 102 TTCCAGGCCAATTGTGGCATAGATTTATCATATTTCTGGATTTTGG 149  
|||||  
RESULT 4  
AC093903/c  
LOCUS AC093903 161280 bp DNA linear PRI 01-MAR-2002  
DEFINITION Homo sapiens BAC clone RP11-733C7 from 4, complete  
ACCESSION AC093903 AC055827  
VERSION AC093903.3 GI:15920156  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Sulston, J.E. and Waterston, R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
PUBMED 9847074  
REFERENCE 2  
AUTHORS Radionenko, M. and Kozlowicz, A.  
TITLE The sequence of Homo sapiens BAC clone RP11-733C7  
JOURNAL Unpublished (2001)  
REFERENCE 3  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (10-SEP-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

## REFERENCE

4 (bases 1 to 161280)  
Waterston,R.H.  
Direct Submission  
Submitted (04-OCT-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

## REFERENCE

5 (bases 1 to 161280)  
Waterston,R.  
Direct Submission  
Submitted (01-MAR-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Oct 4, 2001 this sequence version replaced gi:15625016.

## COMMENT

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0733C07  
Drafting Center: WIBR  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pleter de Jong and coworkers at <http://www.chori.org>  
VECTOR: pBACE3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-310A13. Actual start of this clone is at base position 1 of RP11-733C7; actual end is at base position 161280 of RP11-733C7.

Data from AC009792 was used to finish the clone, AC055827.

The sequence of AC055827 has been incorporated into AC093903.

## FEATURES

## source

Location/Qualifiers  
1. .161280  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-733C7"  
/clone\_lib="RPCI-11"  
148. .304  
/rpt\_family="L1"  
306. .406  
/rpt\_family="(GGAA)n"  
471. .525  
/rpt\_family="AT\_rich"

## repeat\_region

## repeat\_region

## repeat\_region

repeat\_region  
530. .782  
/rpt\_family="L1"  
685. .712  
/rpt\_family="(TA)n"  
1383. .1472  
/rpt\_family="GA-rich"  
1697. .1836  
/rpt\_family="L1"  
1837. .1956  
/rpt\_family="L1"  
3005. .3221  
/rpt\_family="MIR"  
4013. .4321  
/rpt\_family="Alu"  
4295. .4325  
/rpt\_family="(A)n"  
5625. .5960  
/rpt\_family="L1"  
5938. .5982  
/rpt\_family="A-rich"  
6062. .6195  
/rpt\_family="MIR"  
6388. .6462  
/rpt\_family="L2"  
6546. .6868  
/rpt\_family="L1"  
6866. .7318  
/note="similar to EST C06175 (NID:gl502951)"  
6866. .6898  
/rpt\_family="AT\_rich"  
7134. .7385  
/note="similar to EST BG989897 (NID:gl4393967)"  
7777. .7918  
/note="similar to EST BG388107 (NID:gl3281553)"  
8483. .8522  
/rpt\_family="AT\_rich"  
8721. .9018  
/rpt\_family="Alu"  
9459. .9741  
/rpt\_family="L2"  
10217. .10246  
/rpt\_family="AT\_rich"  
10304. .10385  
/rpt\_family="MIR"  
10835. .10878  
/rpt\_family="AT\_rich"  
11861. .11987  
/note="similar to EST BG388107 (NID:gl3281553)"  
11873. .11987  
/note="similar to EST BF286052 (NID:gl1217122)"  
12516. .12539  
/rpt\_family="(T)n"  
12929. .13232  
/rpt\_family="Alu"  
13237. .13274  
/rpt\_family="(TA)n"  
13275. .13304  
/rpt\_family="(T)n"  
13696. .13838  
/rpt\_family="MIR"  
14127. .14153  
/rpt\_family="AT\_rich"  
14132. .14436  
/rpt\_family="Alu"  
15823. .16117  
/rpt\_family="Alu"  
17329. .17894  
/note="similar to EST BG388107 (NID:gl3281553)"  
17329. .17618  
/note="similar to EST BF286052 (NID:gl1217122)"  
17489. .17611  
/note="similar to EST AT313891 (NID:g4029010)"  
17548. .17869  
misc\_feature

misc\_feature /note="similar to EST BF095483 (NID:g10901193)" 17772..18329  
repeat\_region /note="similar to EST BG181259 (NID:g13703031)" 17845..17869  
repeat\_region /rpt\_family="CTG)n" 18792..18828  
repeat\_region /rpt\_family="AT\_rich" 20734..20784  
repeat\_region /rpt\_family="MERL\_type" 20832..20944  
repeat\_region /rpt\_family="MERL\_type" 20948..21964  
repeat\_region /rpt\_family="L1" 21964..21986  
repeat\_region /rpt\_family="AT\_rich" 21965..22276  
repeat\_region /rpt\_family="Alu" 22277..22323  
repeat\_region /rpt\_family="L1" 22341..22681  
repeat\_region /rpt\_family="L1" 22685..22824  
repeat\_region /rpt\_family="Alu" 22805..22824  
repeat\_region /rpt\_family="(A)n" 22826..23575  
repeat\_region /rpt\_family="L1" 23591..23735  
repeat\_region /rpt\_family="MERL\_type" 23735..23735

Alignment Scores:

Pred. No.: 3.31e-05 Length: 161280  
Score: 98.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x AC093903 (1-161280)

Qy 1 PheGlnAlaAsnCysGlyIleAspPheIlePheTrpIlePheTrp 16  
Db 27653 TTCCAGGCCAATGTGGCATAGATTTATCATATCTGGATTGTTGG 27606

RESULT 5  
AF003739  
LOCUS AF003739 41179 bp DNA linear INV 19-APR-2002  
DEFINITION Caenorhabditis elegans cosmid M01D7, complete sequence.  
ACCESSION AF003739  
VERSION AF003739.2 GI:15145732  
KEYWORDS HTG.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans

REFERENCE 1 (bases 1 to 41179) Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis. Waterston, R.  
AUTHORS Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium  
TITLE Science 282 (5396), 2012-2018 (1998)  
JOURNAL 99069613  
MEDLINE 9851916  
PUBMED 9851916

REFERENCE 2 (bases 1 to 41179) Gattung, S. and Goela, D.  
AUTHORS The sequence of C. elegans cosmid M01D7  
TITLE Unpublished (2001)  
JOURNAL 3 (bases 1 to 41179)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (12-MAY-1997) Genome Sequencing Center, Washington University  
REFERENCE 4 (bases 1 to 41179)  
AUTHORS Waterston, R.

TITLE  
JOURNAL

Direct Submission  
Submitted (09-AUG-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA  
5 (bases 1 to 41179)  
Waterston, R.  
Direct Submission  
Submitted (11-OCT-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA  
6 (bases 1 to 41179)  
Waterston, R.  
Direct Submission  
Submitted (19-APR-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA  
On Aug 9, 2001 this sequence version replaced gi:2105482.  
Submitted by:

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: rw@nematode.wustl.edu and jes@sanger.ac.uk

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: rw@nematode.wustl.edu and jes@sanger.ac.uk

COMMENT

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its analysis see:  
(www.wormbase.org/db/seq/sequence?name=M01D7;class=Sequence)

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C53H9, 3300 bp overlap; the 3' cosmid is Y71G12A, 3500 bp overlap. Actual start of this cosmid is at base position 3297 of M01D7; actual end is at 41179 of M01D7.

NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE\_INDEX.html) and The C. elegans ORFeome cloning project (http://wofdb.dfc.harvard.edu/), similarity to other proteins from Blastx analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES  
Source

1. 41179  
/organism="Caenorhabditis elegans"  
/strain="Bristol N2"  
/db\_xref="taxon:6239"  
/chromosome="I"  
/clone="M01D7"  
52. 5004  
/gene="egl-30"  
/note="for a graphical representation of this gene see: {www.wormbase.org/db/seq/sequence?name=M01D7.7a;class=Sequ

gene



CDS  
join(52. .169,2275. .2402,2896. .3113,3180. .3308,3367. .3496, 4211. .4364,4435. .4542,4922. .5004)  
/gene="egl-30"  
/note="C. elegans heterotrimeric G protein (EGL-30) (GB:U56864); contains similarity to Pfam domain: PF00503 (G-alpha), Score=761.7, E-value=1e-225, N=1; coded for by the following C. elegans cDNAs: yk290g12.5, yk329d2.5, yk277b4.5, yk104e6.5, yk176e9.5, yk361c6.5, yk289b6.5, yk343g12.5, yk413b9.5, yk272f9.5, yk462a6.5, yk461h9.5, yk220a11.5, yk215a10.5, yk203e6.5, yk343g12.3, yk622a12.5, yk529e9.5, yk551d12.5, yk654c6.5, yk630b4.5, yk538d9.5, yk587a10.5, yk681a11.5, yk622a12.3, yk654c6.3, yk681e8.3, yk892a08.3, yk1343a05.3, yk551b6.5, yk1247e07.5, yk892a08.5, yk1343a05.5, U56864"  
/codon\_start=1  
/product="C. elegans EGL-30 protein (corresponding sequence M01D7.7a)"  
/protein\_id="AAB58071.1"  
/db\_xref="GI:2105489"  
/translation="MACCLSEAREQKRINOIEKQLQRKRNARRELKLLLTGTGES GKSTFIKQMRIIHGGQYSEEDKRAHIRLVYQNFVMAIQSMIRAMTLDIKFGNESEEL QEKAADVREDFESVTSFEPEYVSYIKELWEDSGIQCYDRREYQLTDSAKYLYSDL RRLAVPDYLPTEQDILRVPTTGIIIEYFPDLEQIIIFRMVDVGGQSRERRKWIHCFEN VTSIMFLVALSEYDQVLVECDNENRMEESKALFRTIITYPWFNTSSVILFLNKKDLE EKILYSHLADYFPEYDGPDPPIAAREFILKMFVDLNPADKTIYSHFTCATDTENIR FVFAAVKDTILQHNLEYNLV"  
1097. .1886  
/gene="M01D7.8"  
/note="for a graphical representation of this gene see: [www.wormbase.org/db/seq/sequence?name=M01D7.8;class=Sequence]"  
join(1097. .1233,1814. .1886)  
/gene="M01D7.8"  
/codon\_start=1  
/product="Hypothetical protein M01D7.8"  
/protein\_id="AAM15594.1"  
/db\_xref="GI:20198866"  
/translation="MEKREEREGESGESHNLSAANQPTNQPAQPTTGRARKSPEKEG WGEFHGLIYLDLQKTREKRRVLN"  
join(2313. .2402,2896. .3113,3180. .3308,3367. .3496, 4211. .4364,4435. .4542,4922. .5004)  
/gene="egl-30"  
/note="C. elegans heterotrimeric G protein (EGL-30); alternatively-spliced form; coded for by the following C. elegans cDNAs: yk514a3.5"  
/codon\_start=1  
/product="C. elegans EGL-30 protein (corresponding sequence M01D7.7b)"  
/protein\_id="AAM15593.1"  
/db\_xref="GI:20198865"  
/translation="MRIIHGGQYSEEDKRAHIRLVYQNFVMAIQSMIRAMTLDIKFG NESEELQEKAAVREVDVESVTSFEPEYVSYIKELWEDSGIQCYDRREYQLTDSAK YYLSDLRRLAVPDYLPTEQDILRVPTTGIIIEYFPDLEQIIIFRMVDVGGQSRERRK W IHCFENVTSIMFLVALSEYDQVLVECDNENRMEESKALFRTIITYPWFNTSSVILFLN KKDLEEKILYSHLADYFPEYDGPDPPIAAREFILKMFVDLNPADKTIYSHFTCAT DTENIRFVFAAVKDTILQHNLEYNLV"  
5560. .6561  
/note="SL1 trans-splice site; see yk589h7.5"  
5576. .7202  
/gene="M01D7.6"  
/note="for a graphical representation of this gene see: [www.wormbase.org/db/seq/sequence?name=M01D7.6;class=Sequence]"  
join(6576. .6714,6789. .7039,7092. .7202)  
/gene="M01D7.6"  
/note="similar to a short region of thymopoietins; coded for by the following C. elegans cDNAs: yk79h4.5, yk113e10.5, yk113e10.3, yk79h4.3, yk142h6.3, yk142h6.5, yk196b5.5, yk258d11.3, yk258d11.5, yk466f7.5, yk589h7.5, yk668d10.5, yk669g2.5, yk669g2.3, yk668d10.3, yk589h7.3"  
/codon\_start=1  
/product="Hypothetical protein M01D7.6"

misc\_feature  
gene  
CDS

/protein\_id="AAB58065.1"  
/db\_xref="GI:2105483"  
/translation="MDVSQLTDAELRDSLKSHGVSVPVATPRKLYEKKLKLSDGSG INNSNLNDSQFNEDSLIISSPKSPQRFVQNVSAATAATSPESDSDDCESMR YLTEEMAADRASARQAQSNKSGFLGSTITFTILFVFIADVFAFLIENAEQLKLV AET NPEDTI"  
7523. .7901  
/gene="nlp-12"  
/note="for a graphical representation of this gene see: [www.wormbase.org/db/seq/sequence?name=M01D7.5;class=Sequence]"  
join(7523. .7577,7657. .7715,7773. .7901)  
/gene="nlp-12"  
/note="C. elegans putative neuropeptide preproprotein (NLP-12); coded for by the following C. elegans cDNAs: yk162h11.3, yk162h11.5, yk276h6.3, yk276h6.5, yk1188b11.3, yk550c1.5"  
/codon\_start=1  
/product="C. elegans NLP-12 protein (corresponding sequence M01D7.5)"  
/protein\_id="AAB58066.1"  
/db\_xref="GI:2105484"  
/translation="MLRHHSCALLMLILVFEVFATQSPTFDRQDRDYRPLQFGKRDG YRPLQFGKRDYRPLQFGKRSSSGSGPVVLEPIWEWQ"  
9337. .14518  
/gene="M01D7.4"  
/note="for a graphical representation of this gene see: [www.wormbase.org/db/seq/sequence?name=M01D7.4;class=Sequence]"

gene  
CDS

Alignment Scores:  
Pred. No.: 2.6 Length: 41179  
Score: 66.00 Matches: 11  
Percent Similarity: 68.75% Conservative: 0  
Best Local Similarity: 68.75% Mismatches: 5  
Query Match: 67.35% Indels: 0  
DB: 3 Gaps: 0

US-09-854-133-587 (1-16) x AF003739 (1-41179)

Qy 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
Db 8533 TTTTCGGCAAATTGTGATTTTGTATTATTATTTTGGAAATTTGG 8580

RESULT 6  
AC006895  
LOCUS AC006895 201746 bp DNA linear HTG 26-FEB-1999  
DEFINITION Caenorhabditis elegans clone Y71G12, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 12 unordered pieces.

ACCESSION AC006895  
VERSION AC006895.2 GI:4309904  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans

REFERENCE  
AUTHORS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE  
AUTHORS 1 (bases 1 to 201746)  
TITLE Waterston,R.H.

JOURNAL  
AUTHORS The sequence of Caenorhabditis elegans clone  
TITLE Unpublished

REFERENCE  
AUTHORS 2 (bases 1 to 201746)  
TITLE Waterston,R.H.

JOURNAL  
AUTHORS Direct Submission  
TITLE Submitted (24-FEB-1999) Genome Sequencing Center, Washington

COMMENT  
AUTHORS University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

On Mar 1, 1999 this sequence version replaced gi:4263453.  
\* NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 2190: contig of 2190 bp in length  
\* 2191 2205: gap of unknown length  
\* 2206 4263: contig of 2058 bp in length  
\* 4264 4278: gap of unknown length  
\* 4279 6646: contig of 2368 bp in length  
\* 6647 6661: gap of unknown length  
\* 6662 10186: contig of 3525 bp in length  
\* 10187 10201: gap of unknown length  
\* 10202 14198: contig of 3997 bp in length  
\* 14199 14213: gap of unknown length  
\* 14214 18232: contig of 4019 bp in length  
\* 18233 18247: gap of unknown length  
\* 18248 23677: contig of 5430 bp in length  
\* 23678 23692: gap of unknown length  
\* 23693 29225: contig of 5533 bp in length  
\* 29226 29240: gap of unknown length  
\* 29241 37388: contig of 8148 bp in length  
\* 37389 37403: gap of unknown length  
\* 37404 69862: contig of 32459 bp in length  
\* 69863 69877: gap of unknown length  
\* 69878 128770: contig of 58893 bp in length  
\* 128771 128785: gap of unknown length  
\* 128786 201746: contig of 72961 bp in length.  
FEATURES  
source  
1. .201746  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/clone="Y71G12"

BASE COUNT 63828 a 37036 c 36424 g 64293 t 165 others  
ORIGIN

Alignment Scores:  
Pred. No.: 15.9 Length: 201746  
Score: 66.00 Matches: 11  
Percent Similarity: 68.75% Conservative: 0  
Best Local Similarity: 68.75% Mismatches: 5  
Query Match: 67.35% Indels: 0  
DB: 2 Gaps: 0

US-09-854-133-587 (1-16) x AC006895 (1-201746)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIlePheTrpIlePheTrp 16  
||| ||||||| ||||||||||| |||||  
Db 84071 TTTCGGCAAATGTGATTTTGTATTATATTTTGGAAATTTGG 84118

RESULT 7  
AX251448  
LOCUS AX251448 24259 bp DNA linear PAT 05-OCT-2001  
DEFINITION Sequence 416 from Patent WO0168912.  
ACCESSION AX251448  
VERSION AX251448.1 GI:15984871  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequences.  
REFERENCE 1 (bases 1 to 24259)  
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.  
TITLE Diagnosis of diseases associated with tumor suppressor genes and oncogenes  
JOURNAL Patent: WO 0168912-A 416 20-SEP-2001;  
Epigenomics AG (DE)  
FEATURES  
source Location/Qualifiers  
1. .24259  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 6444 a 375 c 5691 g 11749 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 10.5 Length: 24259  
Score: 61.00 Matches: 8  
Percent Similarity: 90.91% Conservative: 2

Best Local Similarity: 72.73% Mismatches: 1  
Query Match: 62.24% Indels: 0  
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX251448 (1-24259)

QY 6 GlyIleAspPheIlePheTrpIlePheTrp 16  
||||| ||||||| ::::: ||||| |||||  
Db 21896 GGTATTGATTTTATAGTATTGTTTGGT 21928

RESULT 8  
AC008563/c  
LOCUS AC008563 94703 bp DNA linear HTG 18-JUL-2000  
DEFINITION Homo sapiens chromosome 5 clone CTC-536A23, WORKING DRAFT SEQUENCE,  
7 ordered pieces.  
ACCESSION AC008563  
VERSION AC008563.4 GI:7711299  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL 1 (bases 1 to 94703)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
2 (bases 1 to 94703)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On May 6, 2000 this sequence version replaced gi:7708951.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----

Project Information  
Center Project Name: 391027, H433  
Center clone name: CIT-HSPC\_536A23  
-----  
Summary Statistics  
Consensus quality: 88918 bases at least Q40  
Consensus quality: 93000 bases at least Q30  
Consensus quality: 93760 bases at least Q20  
Estimated insert size: 99000; pulse field gel estimation  
Estimated insert size: 94453; sum-of-contigs estimation  
Quality coverage: 5.84 in Q20 bases; pulse field gel estimation  
Quality coverage: 6.12 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 34594: contig of 34594 bp in length  
\* 34595 34694: gap of unknown length  
\* 34695 51143: contig of 16449 bp in length  
\* 51144 51243: gap of unknown length  
\* 51244 66425: contig of 15182 bp in length  
\* 66426 66525: gap of unknown length  
\* 66526 72366: contig of 5841 bp in length  
\* 72367 72466: gap of unknown length  
\* 72467 79629: contig of 7163 bp in length  
\* 79630 79729: gap of unknown length  
\* 79730 80679: contig of 950 bp in length  
\* 80680 80779: gap of unknown length  
\* 80780 94703: contig of 13924 bp in length.  
FEATURES  
source Location/Qualifiers  
1. .94703

BASE COUNT 29293 a 17932 c 16575 g 30302 t 601 others  
ORIGIN

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTC-536A23"  
/clone\_lib="CalTech human BAC library C"  
29293 a 17932 c 16575 g 30302 t 601 others

Alignment Scores:  
Pred. No.: 74.5 Length: 94703  
Score: 60.00 Matches: 8  
Percent Similarity: 66.67% Conservative: 2  
Best Local Similarity: 53.33% Mismatches: 5  
Query Match: 61.22% Indels: 0  
DB: 2 Gaps: 0

US-09-854-133-587 (1-16) x AC008563 (1-94703)

QY 2 GlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
Db 79322 CAAAGGAAATGCTCATTAAGAAATTTTTCGATTTGGATTTTGG 79278

RESULT 9  
AC016603  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone CTD-2132G23, complete sequence.  
AC016603  
VERSION AC016603.6 GI:14971185  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 110879) DNA linear PRI 20-JUL-2001  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL  
REFERENCE 2 (bases 1 to 110879)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL  
REFERENCE 3 (bases 1 to 110879)  
AUTHORS Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint  
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
JOURNAL  
REFERENCE 3 (bases 1 to 110879)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL  
REFERENCE 3 (bases 1 to 110879)  
AUTHORS Submitted (20-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell  
TITLE Drive, Walnut Creek, CA 94598, USA  
JOURNAL  
COMMENT On Jul 20, 2001 this sequence version replaced gi:7711572.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www-shgc.stanford.edu  
Quality: Phrap Quality >=40 99.9% of Sequence;  
Estimated Total Number of Errors is 0.1.  
STS Content:  
WI-31186 G24491.

FEATURES  
source  
1. .110879  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2132G23"  
33303 a 18836 c 20919 g 37821 t

Alignment Scores:  
Pred. No.: 89.2 Length: 110879  
Score: 60.00 Matches: 8  
Percent Similarity: 66.67% Conservative: 2  
Best Local Similarity: 53.33% Mismatches: 5  
Query Match: 61.22% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x AC016603 (1-110879)  
QY 2 GlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
Db 58777 CAAAGGAAATGCTCATTAAGAAATTTTTCGATTTGGATTTTGG 58821

RESULT 10  
AC010359/c  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone CTD-2033C11, complete sequence.  
AC010359  
VERSION AC010359.5 GI:11968292  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 204843) DNA linear PRI 22-DEC-2000  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL  
REFERENCE 2 (bases 1 to 204843)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL  
REFERENCE 2 (bases 1 to 204843)  
AUTHORS Unpublished  
TITLE Direct Submission  
JOURNAL  
REFERENCE 2 (bases 1 to 204843)  
AUTHORS DOE Joint Genome Institute.

REFERENCE 3 (bases 1 to 204843)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL  
REFERENCE 3 (bases 1 to 204843)  
AUTHORS Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint  
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
JOURNAL  
REFERENCE 3 (bases 1 to 204843)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL  
COMMENT On Dec 22, 2000 this sequence version replaced gi:7710748.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www-shgc.stanford.edu  
Quality: Phrap Quality >=40 99.7% of Sequence;  
Estimated Total Number of Errors is 0.5.  
STS Content:  
WI-31186 G24491.

FEATURES  
source  
1. .204843  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2033C11"  
66824 a 40891 c 37481 g 59647 t

Alignment Scores:  
Pred. No.: 180 Length: 204843  
Score: 60.00 Matches: 8  
Percent Similarity: 66.67% Conservative: 2  
Best Local Similarity: 53.33% Mismatches: 5  
Query Match: 61.22% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x AC010359 (1-204843)

QY 2 GlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
Db 37059 CAAAGGAAATGCTCATTAAGAAATTTTTCGATTTGGATTTTGG 37015

RESULT 11  
AC113545/c  
LOCUS  
DEFINITION Mus musculus clone RP23-241C6, LOW-PASS SEQUENCE SAMPLING.  
AC113545  
VERSION AC113545.1 GI:19033729  
KEYWORDS HTG; HTGS\_PHASE0.  
SOURCE Mus musculus.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 67679)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Mus musculus, clone RP23-241C6  
Unpublished

2 (bases 1 to 67679)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,  
Ferreira, P., FlitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Illiev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,  
Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,  
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,  
McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T.,  
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,  
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT

Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L24237

Center clone name: 241\_C\_6

-----

\* NOTE: This record contains 82 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
704 803: contig of 703 bp in length  
804 803: gap of 100 bp  
1496 1595: contig of 692 bp in length  
1596 2332: contig of 737 bp in length  
2333 2432: gap of 100 bp  
2433 3158: contig of 726 bp in length  
3159 3258: gap of 100 bp  
3259 3968: contig of 710 bp in length  
3969 4068: gap of 100 bp  
4069 4760: contig of 692 bp in length  
4761 4860: gap of 100 bp  
4861 5578: contig of 718 bp in length  
5579 5678: gap of 100 bp  
5679 6403: contig of 725 bp in length  
6404 6503: gap of 100 bp  
6504 7236: contig of 733 bp in length  
7237 7336: gap of 100 bp  
7337 8074: contig of 738 bp in length

8075 8174: gap of 100 bp  
8175 8880: contig of 706 bp in length  
8881 8980: gap of 100 bp  
8981 9708: contig of 728 bp in length  
9709 9808: gap of 100 bp  
9809 10544: contig of 736 bp in length  
10545 10644: gap of 100 bp  
10645 11350: contig of 706 bp in length  
11351 11450: gap of 100 bp  
11451 12165: contig of 715 bp in length  
12166 12265: gap of 100 bp  
12266 12984: contig of 719 bp in length  
12985 13084: gap of 100 bp  
13085 13817: contig of 733 bp in length  
13818 13917: gap of 100 bp  
13918 14663: contig of 746 bp in length  
14664 14763: gap of 100 bp  
14764 15513: contig of 750 bp in length  
15514 15613: gap of 100 bp  
15614 16344: contig of 731 bp in length  
16345 16444: gap of 100 bp  
16445 17153: contig of 709 bp in length  
17154 17253: gap of 100 bp  
17254 17993: contig of 740 bp in length  
17994 18093: gap of 100 bp  
18094 18837: contig of 744 bp in length  
18838 18937: gap of 100 bp  
18938 19685: contig of 748 bp in length  
19686 19785: gap of 100 bp  
19786 20501: contig of 716 bp in length  
20502 20601: gap of 100 bp  
20602 21311: contig of 710 bp in length  
21312 21411: gap of 100 bp  
21412 22130: contig of 719 bp in length  
22131 22230: gap of 100 bp  
22231 22944: contig of 714 bp in length  
22945 23044: gap of 100 bp  
23045 23782: contig of 738 bp in length  
23783 23882: gap of 100 bp  
23883 24623: contig of 741 bp in length  
24624 24723: gap of 100 bp  
24724 25469: contig of 746 bp in length  
25470 25569: gap of 100 bp  
25570 26275: contig of 706 bp in length  
26276 26375: gap of 100 bp  
26376 27120: contig of 745 bp in length  
27121 27220: gap of 100 bp  
27221 27906: contig of 686 bp in length  
27907 28006: gap of 100 bp  
28007 28732: contig of 726 bp in length  
28733 28832: gap of 100 bp  
28833 29544: contig of 712 bp in length  
29545 29644: gap of 100 bp  
29645 30372: contig of 728 bp in length  
30373 30472: gap of 100 bp  
30473 31190: contig of 718 bp in length  
31191 31290: gap of 100 bp  
31291 32018: contig of 728 bp in length  
32019 32118: gap of 100 bp  
32119 32871: contig of 753 bp in length  
32872 32971: gap of 100 bp  
32972 33702: contig of 731 bp in length  
33703 33802: gap of 100 bp  
33803 34551: contig of 749 bp in length  
34552 34651: gap of 100 bp  
34652 35355: contig of 704 bp in length  
35356 35455: gap of 100 bp  
35456 36199: contig of 744 bp in length  
36200 36299: gap of 100 bp  
36300 37014: contig of 715 bp in length  
37015 37114: gap of 100 bp  
37115 37824: contig of 710 bp in length  
37825 37924: gap of 100 bp



\* 37925 38637: contig of 713 bp in length  
\* 38638 38737: gap of 100 bp  
\* 38738 39421: contig of 684 bp in length  
\* 39422 39521: gap of 100 bp  
\* 39522 40255: contig of 734 bp in length  
\* 40256 40355: gap of 100 bp  
\* 40356 41119: contig of 764 bp in length  
\* 41120 41219: gap of 100 bp  
\* 41220 41933: contig of 714 bp in length  
\* 41934 42033: gap of 100 bp  
\* 42034 42766: contig of 733 bp in length  
\* 42767 42866: gap of 100 bp  
\* 42867 43599: contig of 733 bp in length  
\* 43600 43699: gap of 100 bp  
\* 44428 44428: contig of 729 bp in length  
\* 44429 44528: gap of 100 bp  
\* 44529 45245: contig of 717 bp in length  
\* 45246 45345: gap of 100 bp  
\* 45346 46066: contig of 721 bp in length  
\* 46067 46166: gap of 100 bp  
\* 46167 46889: contig of 723 bp in length  
\* 46890 46989: gap of 100 bp  
\* 46990 47715: contig of 726 bp in length  
\* 47716 47815: gap of 100 bp  
\* 47816 48569: contig of 754 bp in length  
\* 48570 48669: gap of 100 bp  
\* 48670 49396: contig of 727 bp in length  
\* 49397 49496: gap of 100 bp  
\* 49497 50223: contig of 727 bp in length  
\* 50224 50323: gap of 100 bp  
\* 50324 51034: contig of 711 bp in length  
\* 51035 51134: gap of 100 bp  
\* 51135 51883: contig of 749 bp in length  
\* 51884 51983: gap of 100 bp  
\* 51984 52732: contig of 749 bp in length  
\* 52733 52832: gap of 100 bp  
\* 52833 53558: contig of 726 bp in length  
\* 53559 53658: gap of 100 bp  
\* 53659 54380: contig of 722 bp in length  
\* 54381 54480: gap of 100 bp  
\* 54481 55198: contig of 718 bp in length  
\* 55199 55298: gap of 100 bp  
\* 55299 56015: contig of 717 bp in length  
\* 56016 56115: gap of 100 bp  
\* 56116 56842: contig of 727 bp in length  
\* 56843 56942: gap of 100 bp

Alignment Scores:

Pred. No.: 75.8 Length: 67679  
Score: 59.00 Matches: 7  
Percent Similarity: 90.91% Conservative: 3  
Best Local Similarity: 63.64% Mismatches: 1  
Query Match: 60.20% Indels: 0  
DB: 2 Gaps: 0

US-09-854-133-587 (1-16) x AC113545 (1-67679)

QY 6 GlyIleAspPheIleIlePheTrpIlePheTrp 16

Db 26971 GGCTGGATTGTTTGGTTTGGTTTGGTTTGG 26939

RESULT 12

AC091629 161334 bp DNA linear ROD 19-DEC-2001  
LOCUS Mus musculus chromosome 11 map 11, 64cm, Band E1, complete  
DEFINITION sequence.

AC091629 AC091629.1 GI:14029058

VERSION HTG.

KEYWORDS Mus musculus.

SOURCE Mus musculus.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS

1 (bases 1 to 161334)  
Poorkaj,P., Kas,A., D'Souza,I., Zhou,Y., Pham,Q., Stone,M.,

Olson,M.V. and Schellenberg,G.D.

A genomic sequence analysis of the mouse and human

microtubule-associated protein tau

Mamm. Genome 12 (9), 700-712 (2001)

21521499

11641718

Contact: Gerald D. Schellenberg (zachdad@u.washington.edu)

2 (bases 1 to 161334)

Kas,A.

Direct Submission

Submitted (12-MAY-2001) Human Genome Center, University of

Washington, Box 352145, Seattle, WA 98195, USA

Contact: Arnold Kas(akas@u.washington.edu)

Sequence Validation:

This sequence has been validated by Multiple Complete Digest

fingerprinting. Comparison of the experimentally derived digest

fragments with sequence-predicted fragments is given below

(only fragments from insert sequence is displayed). Small

fragments below a variable cutoff (approximately 400-600 bp)

are not resolved in the fingerprint and hence do not appear

in the table. There are no significant remaining discrepancies

between the experimental and predicted values. Uniquely ordered

fragments are separated by dashed lines.

BglII EcoRI

FP Seq FP Seq

1991.00 2013.00 2272.00 2258.00 4046.00 4032.00

1788.00 1820.00 2854.00 2843.00 6952.00 6857.00

3992.00 4005.00 9569.00 9597.00 506.00 508.00

2149.00 2138.00 2099.00 2105.00 5420.00 5320.00

1788.00 1753.00 1599.00 1581.00 1242.00 1233.00

1879.00 1869.00 7895.00 7867.00 25755.00 22305.00

1226.00 1216.00 8085.00 7933.00 7924.00 7633.00

5423.00 5449.00 18024.00 17543.00 13876.00 13549.00

7084.00 6935.00 828.00 801.00 7361.00 7144.00

2654.00 2591.00 4216.00 4115.00 3148.00 3089.00

2084.00 2056.00 5017.00 4935.00 2756.00 2736.00

2654.00 2586.00 11890.00 11718.00 7924.00 7789.00

985.00 970.00 3794.00 3736.00 5248.00 5145.00

9434.00 9229.00 5245.00 5170.00 3324.00 3278.00

869.00 851.00 4608.00 4523.00 6952.00 6725.00

4250.00 4176.00 4216.00 4164.00 3823.00 3755.00

1100.00 1081.00 1898.00 1875.00 1666.00 1678.00

1541.00 1525.00 2828.00 2786.00 6602.00 6496.00

4250.00 4230.00 625.00 613.00 10116.00 9829.00

3992.00 3884.00 9258.00 9151.00 9151.00 8941.00

3544.00 3448.00 1453.00 1455.00 3324.00 3257.00

1226.00 1198.00 2416.00 2363.00

1991.00 1950.00 1311.00 1307.00

5963.00	5822.00	4331.00	4286.00
2084.00	2107.00	1453.00	1394.00
869.00	850.00	3208.00	3152.00
1610.00	1594.00	1102.00	1090.00
1461.00	1442.00	2334.00	2332.00
2811.00	2734.00	1898.00	1900.00
573.00	566.00	14838.00	14692.00
4559.00	4489.00	6683.00	6584.00
3208.00	3170.00	1453.00	1443.00
1788.00	1723.00	3208.00	3153.00
1461.00	1452.00	916.00	899.00
4559.00	4483.00	2957.00	2900.00
2254.00	2208.00		
1788.00	1789.00		
7270.00	7150.00		
4834.00	4757.00		
3992.00	3956.00		
1461.00	1474.00		
1395.00	1392.00		
2654.00	2587.00		
16421.00	16316.00		
1879.00	1857.00		
6214.00	6135.00		

FEATURES source Location/Qualifiers  
1. .161334  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/map="11, 64cM, Band E1"  
/cell\_line="Embryonic stem cell 129/svj"  
/clone\_lib="Genome Systems ES Release I"  
BASE COUNT 42340 a 37230 c 38866 g 42898 t  
ORIGIN

Alignment Scores:  
Pred. No.: 204 Length: 161334  
Score: 59.00 Matches: 7  
Percent Similarity: 90.91% Conservative: 3  
Best Local Similarity: 63.64% Mismatches: 1  
Query Match: 60.20% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x AC091629 (1-161334)  
QY 6 GlyIleAspPheIleIlePheTrpIlePheTrp 16  
||||:|||||:||||:|||||  
Db 36533 GGCTTGGATTTTGTGTTTGGTTTGGT 36565  
RESULT 13

AL593843  
LOCUS  
DEFINITION Mouse DNA sequence from clone RP23-136D4 on chromosome 11, complete sequence.  
ACCESSION AL593843  
VERSION AL593843.9 GI:21211822  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 198281)  
Peck, A.  
Direct Submission  
Submitted (22-MAY-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
On May 25, 2002 this sequence version replaced gi:16944181.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep RP23-136D4 is from the RPI-23 Mouse PAC Library  
constructed by the group of Pieter de Jong.  
For further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pBACE3.6.

FEATURES source Location/Qualifiers  
1. .198281  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/clone="RP23-136D4"  
/clone\_lib="RPI-23"  
BASE COUNT 53943 a 41053 c 44095 g 59190 t  
ORIGIN

Alignment Scores:  
Pred. No.: 259 Length: 198281  
Score: 59.00 Matches: 7  
Percent Similarity: 90.91% Conservative: 3  
Best Local Similarity: 63.64% Mismatches: 1  
Query Match: 60.20% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x AL593843 (1-198281)  
QY 6 GlyIleAspPheIleIlePheTrpIlePheTrp 16  
||||:|||||:||||:|||||  
Db 130121 GGCTTGGATTTTGTGTTTGGTTTGGT 130153  
RESULT 14

AL845270/c  
LOCUS  
DEFINITION Mus musculus chromosome 2 clone RP23-186C16, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
ACCESSION AL845270  
VERSION AL845270.4 GI:22416234  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE house mouse.



```
repeat_region 1159. .1367 /rpt_family="MERL_type"
repeat_region 1522. .2141 /rpt_family="MERL_type"
repeat_region 4257. .4555 /rpt_family="Alu"
repeat_region 5498. .5793 /rpt_family="Alu"
repeat_region 5976. .6010 /rpt_family="MIR"
repeat_region 7559. .7838 /rpt_family="Retroviral"
repeat_region 7869. .7908 /rpt_family="Retroviral"
repeat_region 7909. .8211 /rpt_family="Alu"
repeat_region 8212. .8574 /rpt_family="Retroviral"
repeat_region 8851. .9024 /rpt_family="MERL_type"
repeat_region 9084. .9288 /rpt_family="MERL_type"
repeat_region 9138. .9295 /rpt_family="MERL_type"
repeat_region 9303. .9383 /rpt_family="MERL_type"
repeat_region 9582. .10468 /rpt_family="MERL_type"
repeat_region 10469. .10727 /rpt_family="MERL_type"
repeat_region 11285. .11446 /rpt_family="L2"
repeat_region 11556. .11616 /rpt_family="MERL_type?"
repeat_region 11933. .12100 /rpt_family="MERL_type"
repeat_region 12623. .12969 /rpt_family="MaLR"
repeat_region 13221. .13348 /rpt_family="L2"
repeat_region 13949. .14093 /rpt_family="L2"
repeat_region 14651. .14867 /rpt_family="L2"
repeat_region 14943. .15069 /rpt_family="L2"
repeat_region 15381. .15518 /rpt_family="L2"
repeat_region 17545. .17685 /rpt_family="MIR"
repeat_region 17719. .17994 /rpt_family="Alu"
repeat_region 18973. .19281 /rpt_family="Alu"
repeat_region 20216. .20342 /rpt_family="MIR"
repeat_region 20478. .20748 /rpt_family="MERL_type"
repeat_region 20817. .21122 /rpt_family="Alu"
repeat_region 21183. .21465 /rpt_family="MERL_type"
repeat_region 21466. .21769 /rpt_family="Alu"
repeat_region 21786. .22056 /rpt_family="MaLR"
repeat_region 22144. .22284 /rpt_family="MERL_type"
repeat_region 22294. .22546 /rpt_family="MERL_type"
repeat_region 26976. .27281 /rpt_family="Alu"
repeat_region 29024. .29089 /rpt_family="Alu"

repeat_region 1000000. .1000000 /rpt_family="MIR"
repeat_region 1000000. .1000000 /rpt_family="Alu"
repeat_region 1000000. .1000000 /rpt_family="L1"
repeat_region 1000000. .1000000 /rpt_family="L1"
STS complement(36940. .37258)
/db_xref="GI:1185354"
37403. .37491 /rpt_family="L2"
38538. .38572 /rpt_family="MIR"
40350. .40770 /rpt_family="MaLR"
41667. .41912 /rpt_family="MIR"
41902. .41973 /rpt_family="MIR"
43348. .43391 /rpt_family="MaLR"
43455. .43692 /rpt_family="MaLR"
44189. .44400 /rpt_family="L1"
44397. .44606 /rpt_family="L1"
44607. .44892 /rpt_family="L1"
44893. .45706 /rpt_family="L1"
45707. .46031 /rpt_family="Alu"
46032. .46161 /rpt_family="L1"
46162. .46469 /rpt_family="Alu"
46470. .47221 /rpt_family="L1"
47222. .47523 /rpt_family="Alu"
47524. .47745 /rpt_family="Alu"

Alignment Scores:
Pred. No.: 419
Score: 57.00
Percent Similarity: 83.33%
Best Local Similarity: 58.33%
Query Match: 58.16%
DB: 9
Length: 149765
Matches: 7
Conservative: 3
Mismatch: 2
Indels: 0
Gaps: 0

US-09-854-133-587 (1-16) x AC004984 (1-149765)
QY 5 CysGlyIleAspPheIleIlePheTrpIlePheTrp 16
Db 58017 TGCTCTATGGATGCCATCTTTCTGGGCTCTTTGG 57982

Search completed: May 11, 2003, 15:01:07
Job time : 461.761 secs
```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 12:41:29 ; Search time 34.6903 Seconds  
(without alignments)  
1038.677 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 98  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_173955\_327/app\_query.fasta\_1.462  
-DB=N\_Geneseq\_101002 -QFWT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptp -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn\_1\_1\_275\_@runat\_05052003\_173955\_327 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	98	100.0	337	23	AAD23462	Human lung tumour-
2	98	100.0	2239	23	AAD23460	Human lung tumour-
3	98	100.0	5981	23	AAD23461	Human lung tumour-
4	61	62.2	24259	22	AAS46692	Tumour suppressor
5	56	57.1	31882	23	ABL09072	Drosophila melanog
6	55.5	56.6	6526	24	ABN80307	Human chemically m
7	55.5	56.6	6526	24	ABL34237	Human immune syste
8	55	56.1	5575	24	AAS61206	Human gene regulat
9	53	54.1	97835	24	ABK84796	Human CDNA differe
10	52	53.1	740	22	AAH73248	Human cervical can
11	52	53.1	3171	23	ABL19819	Drosophila melanog
12	52	53.1	3430	23	ABL16345	Drosophila melanog
13	52	53.1	6089	24	ABL32703	Human immune syste
14	52	53.1	6285	24	ABL33497	Human immune syste
15	52	53.1	8164	23	ABL19818	Drosophila melanog
16	52	53.1	11322	23	ABL16344	Drosophila melanog
17	52	53.1	611590	21	AAF22303	Arabidopsis thalia
18	51	52.0	10957	24	ABL33111	Human immune syste
19	51	52.0	24939	24	ABL70569	Chemically treated
20	50.5	51.5	180	8	AAH70765	Partial sequence o
21	50.5	51.5	6713	22	AAK84891	Human immune/haema
22	50	51.0	2000	22	AAS32537	Human genomic DNA
23	50	51.0	2584	19	AAV40502	Human secreted pro
24	50	51.0	8889	24	ABL32210	Human immune syste
25	50	51.0	8925	22	AAI07244	Human reproductive
26	50	51.0	8925	23	ABL98792	Human testicular a
27	50	51.0	14245	21	AAZ45259	DNA encoding enzym
28	50	51.0	14245	21	AAZ30356	Nucleotide sequenc
29	50	51.0	15416	24	ABL70498	Chemically treated
30	50	51.0	15416	24	ABL34231	Human immune syste
31	50	51.0	15416	24	AAS61453	Human gene regulat
32	50	51.0	15416	24	ABK31527	Signal transductio
33	50	51.0	15732	22	AAS45389	Chemically pretrea
34	50	51.0	15732	24	ABK28234	DNA transcription
35	50	51.0	40324	24	ABQ67149	Human angiogenesis
36	49	50.0	344	22	AAI81960	Human polynucleoti
37	49	50.0	582	24	ABQ26448	Oligonucleotide fo
38	49	50.0	582	24	ABQ26449	Oligonucleotide fo
39	49	50.0	583	24	ABQ22078	Oligonucleotide fo
40	49	50.0	583	24	ABQ22079	Oligonucleotide fo
41	49	50.0	4579	23	ABL05382	Drosophila melanog
42	49	50.0	5143	24	ABL34567	Human metastasis a
43	49	50.0	5659	24	ABL32797	Human immune syste
44	49	50.0	6113	24	ABL32802	Human immune syste
45	49	50.0	17389	24	ABL33415	Human immune syste

ALIGNMENTS

RESULT 1  
AAD23462  
ID AAD23462 standard; cDNA; 337 BP.  
XX  
AC AAD23462;  
XX 26-FEB-2002 (first entry)  
XX Human lung tumour-specific 20E10 5' cDNA.  
DE Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
KW antisense-therapy; vaccine; immune response; lung cancer; 20E10; ss.  
XX Homo sapiens.  
OS  
XX WO200172295-A2.  
PN  
PD 04-OCT-2001.

```
XX PF 28-MAR-2001; 2001WO-US09991.
XX PR
XX PR 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX PA (CORI-) CORIXA CORP.
XX PI
XX PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX DR WPI; 2001-639201/73.
XX PT
XX PT New human lung-specific polynucleotides and polypeptides for the
XX diagnosis and treatment of disease e.g. lung cancer -
XX PS Claim 1; Page 334; 378pp; English.
XX CC The invention relates to isolated lung tumour-specific proteins and
XX their corresponding cDNA molecules. Lung tumour-specific proteins and
XX their antigen-presenting cells are useful for stimulating and/or
XX expanding T cells specific for a tumour protein, and for inhibiting
XX the development of cancer. The invention also relates to a composition
XX useful for stimulating an immune response, and for treating cancer. The
XX lung tumour specific oligonucleotide is useful in gene therapy and for
XX diagnosis, detection and treatment of lung cancer. The present sequence
XX is a cDNA encoding human lung tumour-specific protein.
XX SQ Sequence 337 BP; 103 A; 60 C; 93 G; 81 T; 0 other;
Alignment Scores:
Pred. No.: 1.67e-06 Length: 337
Score: 98.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0
US-09-854-133-587 (1-16) x AAD23462 (1-337)
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
Db 107 TTCCAGGCCAATTGTGGCATAGATTTATCATATTTCTGGATTTTGG 154
RESULT 2
AAD23460
ID AAD23460 standard; cDNA; 2239 BP.
XX AC AAD23460;
XX XX
XX DT 26-FEB-2002 (first entry)
XX DE Human lung tumour-specific 19A4 cDNA.
XX KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
XX KW antisense-therapy; vaccine; immune response; lung cancer; 19A4; ss.
XX OS Homo sapiens.
XX PN WO200172295-A2.
XX PD 04-OCT-2001.
XX PF 28-MAR-2001; 2001WO-US09991.
XX PR 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
XX PI
```

```
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX PA (CORI-) CORIXA CORP.
XX PI
XX PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX DR WPI; 2001-639201/73.
XX PT
XX PT New human lung-specific polynucleotides and polypeptides for the
XX diagnosis and treatment of disease e.g. lung cancer -
XX PS Claim 1; Page 332; 378pp; English.
XX CC The invention relates to isolated lung tumour-specific proteins and
XX their corresponding cDNA molecules. Lung tumour-specific proteins and
XX their antigen-presenting cells are useful for stimulating and/or
XX expanding T cells specific for a tumour protein, and for inhibiting
XX the development of cancer. The invention also relates to a composition
XX useful for stimulating an immune response, and for treating cancer. The
XX lung tumour specific oligonucleotide is useful in gene therapy and for
XX diagnosis, detection and treatment of lung cancer. The present sequence
XX is a cDNA encoding human lung tumour-specific protein.
XX SQ Sequence 2239 BP; 619 A; 444 C; 493 G; 683 T; 0 other;
Alignment Scores:
Pred. No.: 1.32e-05 Length: 2239
Score: 98.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0
US-09-854-133-587 (1-16) x AAD23460 (1-2239)
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
Db 104 TTCCAGGCCAATTGTGGCATAGATTTATCATATTTCTGGATTTTGG 151
RESULT 3
AAD23461
ID AAD23461 standard; cDNA; 5981 BP.
XX AC AAD23461;
XX XX
XX DT 26-FEB-2002 (first entry)
XX DE Human lung tumour-specific 14F10 full length cDNA.
XX KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
XX KW antisense-therapy; vaccine; immune response; lung cancer; 14F10; ss.
XX OS Homo sapiens.
XX PN WO200172295-A2.
XX PD 04-OCT-2001.
XX PF 28-MAR-2001; 2001WO-US09991.
XX PR 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX PA (CORI-) CORIXA CORP.
XX PI
XX PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
```

XX WPI; 2001-639201/73.  
XX New human lung-specific polynucleotides and polypeptides for the  
PT diagnosis and treatment of disease e.g. lung cancer -  
XX  
XX Claim 1; Page 332-334; 378pp; English.  
XX  
XX The invention relates to isolated lung tumour-specific proteins and  
CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
CC their antigen-presenting cells are useful for stimulating and/or  
CC expanding T cells specific for a tumour protein, and for inhibiting  
CC the development of cancer. The invention also relates to a composition  
CC useful for stimulating an immune response, and for treating cancer. The  
CC lung tumour specific oligonucleotide is useful in gene therapy and for  
CC diagnosis, detection and treatment of lung cancer. The present sequence  
CC is a cDNA encoding human lung tumour-specific protein.  
XX  
XX Sequence 5981 BP; 1858 A; 1029 C; 1098 G; 1996 T; 0 other;  
SQ  
  
Alignment Scores:  
Pred. No.: 3.88e-05 Length: 5981  
Score: 98.00 Matches: 16  
Percent Similarity: 100.00% Conservativity: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 23 Gaps: 0  
  
US-09-854-133-587 (1-16) x AAD23461 (1-5981)  
  
Qy 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
Db 102 TTCCAGGCCAATTGTGGCATAGATTTTATCATATCTGTGGATTTTGG 149  
  
RESULT 4  
AAS46692  
ID AAS46692 standard; DNA; 24259 BP.  
XX  
AC AAS46692;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Tumour suppressor gene derived chemically modified sequence #415.  
XX  
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;  
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
KW cytosine methylation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200168912-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 15-MAR-2001; 2001WO-EP02955.  
XX  
PR 15-MAR-2000; 2000DE-1013847.  
PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2001-602752/68.  
XX  
XX Fragments of chemically modified genes associated with tumour suppressor  
PT genes and oncogenes, useful in designing primers and probes for  
PT analysing diseases associated with cytosine methylation state e.g.  
PT cancer -  
XX

PS Claim 1; SEQ ID No 415; 27pp; English.  
XX  
CC The invention relates to a nucleic acid comprising a sequence of 18  
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
CC bisulphite, of genes associated with tumour suppression and  
CC oncogenes having a sequence taken from 536 (actually 533 since  
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
CC (SS) and sequences complementary to (SS). The nucleic acid may be a  
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
CC form part of a set of probes for detecting the cytosine methylation state  
CC and/or single nucleotide polymorphisms and also to be used in an  
CC array for analysing diseases associated with CpG dinucleotides e.g.  
CC cancers and tumours. The probes can also be used in a method for  
CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
CC and/or therapy of existing diseases or the predisposition to specific  
CC diseases, by analysing cytosine methylations. The parameters may be  
CC compared to another set of genetic and/or epigenetic parameters, the  
CC differences serving as basis for diagnosis and/or prognosis events which  
CC are disadvantageous to patients. The present sequence is one of the  
CC 533 genomic sequences derived from tumour suppressor genes and  
CC oncogenes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 24259 BP; 6444 A; 375 C; 5691 G; 11749 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 88.5 Length: 24259  
Score: 61.00 Matches: 8  
Percent Similarity: 90.91% Conservativity: 2  
Best Local Similarity: 72.73% Mismatches: 1  
Query Match: 62.24% Indels: 0  
DB: 22 Gaps: 0  
  
US-09-854-133-587 (1-16) x AAS46692 (1-24259)  
  
Qy 6 GlyIleAspPheIleIlePheTrpIlePheTrp 16  
Db 21896 GGTATTGATTTTATGATTTTGGTTTGGT 21928  
  
RESULT 5  
ABL09072/c  
ID ABL09072 standard; cDNA; 31882 BP.  
XX  
AC ABL09072;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 21698.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR P-PSDB; ABB64969.  
DR  
XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

PS Claim 1; SEQ ID NO 21698; 21pp + Sequence Listing; English.

XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 31882 BP; 9206 A; 6377 C; 6605 G; 9694 T; 0 other;

Alignment Scores:  
Pred. No.: 701 Length: 31882  
Score: 56.00 Matches: 7  
Percent Similarity: 76.92% Conservative: 3  
Best Local Similarity: 53.85% Mismatches: 3  
Query Match: 57.14% Indels: 0  
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x ABL09072 (1-31882)

QY 4 AsnCysGlyIleAspPheIlePheTrpIlePheTrp 16

Db 27609 AATTGGGGTACGTTTGTGTTTCCCTTGGCTATTTGG 27571

RESULT 6

ABN80307

ID ABN80307 standard; DNA; 6526 BP.

XX

AC ABN80307;

DT 15-JUL-2002 (first entry)

XX

DE Human chemically modified disease associated gene SEQ ID NO 324.

XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
KW antidiabetic; cytostatic; anticonvulsant; ds.

XX Homo sapiens.

OS Synthetic.

XX WO200200927-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07536.

XX 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130908/17.

XX Novel nucleic acid useful for diagnosis and therapy of diseases

XX associated with development genes such as diabetes, comprises a

XX sequence of a segment of chemically pretreated DNA of genes associated

XX with development

XX Claim 1; SEQ ID NO 324; 27pp; English.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX The invention relates to a nucleic acid (I) comprising a sequence at  
CC least 18 bases in length of a segment of chemically pretreated DNA (II)  
CC of genes associated with development selected from 87 genes listed in  
CC the specification such as ACCPN, ADFN, or AFD1 and comprising one of 350  
CC sequences (ABN79984-ABN80333) or their complements. The invention is  
CC useful for the diagnosis or therapy of diseases associated with  
CC development genes, in particular disease related to homeobox containing  
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
CC associated with congenital heart disease, epilepsy, diseases related to  
CC histone deacetylation, Currrarino syndrome, diseases related with the  
CC development of the brain and limb girdle muscular dystrophy and dwarfism.  
CC Oligomers specific to each of the genes are useful for detecting the  
CC methylation state of all CPG dinucleotides within the 350 sequences or  
CC (II) and their complementary sequences, as primer oligonucleotides for  
CC the amplification of the 350 sequences, (II) and/or their complements and  
CC as oligomer probes for detecting the cytosine methylation state and/or  
CC single nucleotide polymorphisms (SNPs).

CC Note: The sequence data for this patent did not form part of the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office.

XX  
SQ Sequence 6526 BP; 1009 A; 556 C; 2240 G; 2721 T; 0 other;

Alignment Scores:

Pred. No.: 148 Length: 6526  
Score: 55.50 Matches: 9  
Percent Similarity: 73.33% Conservative: 2  
Best Local Similarity: 60.00% Mismatches: 1  
Query Match: 56.63% Indels: 3  
DB: 24 Gaps: 1

US-09-854-133-587 (1-16) x ABN80307 (1-6526)

QY 5 CysGlyIleAspPheIlePheTrp-----IlePheTrp 16

Db 1151 TGTGGAATTGCTTTTATAGTTTTTGGTTTTAGCGTTATTTGG 1195

RESULT 7

ABL34237

ID ABL34237 standard; DNA; 6526 BP.

XX

AC ABL34237;

XX

DT 26-MAR-2002 (first entry)

XX

DE Human immune system associated gene SEQ ID NO: 2210.

XX

KW Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.

XX Homo sapiens.

OS WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX



XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation  
XX  
PS Claim 1; SEQ ID NO 2210; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 6526 BP; 1009 A; 556 C; 2240 G; 2721 T; 0 other;  
XX  
Alignment Scores: 148 Length: 6526  
Pred. No.: 55.50 Matches: 9  
Score: 73.33% Conservative: 2  
Percent Similarity: 60.00% Mismatches: 1  
Best Local Similarity: 56.63% Indels: 3  
Query Match: 24 Gaps: 1  
DB: 1  
US-09-854-133-587 (1-16) x ABL34237 (1-6526)  
QY 5 CysGlyIleAspPheIleIlePheTrp-----IlePheTrp 16  
Db 1151 TGTGGAATTGGTTTATAGTTTGTGTTTGGTTTATGCGTTATTTGG 1195  
RESULT 8  
AAS61206  
ID AAS61206 standard; DNA; 5575 BP.  
XX  
AC AAS61206;  
XX  
DT 29-JAN-2002 (first entry)  
XX  
DE Human gene regulation-associated gene oligonucleotide #161.  
XX  
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;  
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;  
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;  
KW renal disease; Preeclampsia; cardiac allograft vascular disease;  
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;  
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;  
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.  
XX  
OS Homo sapiens.  
XX  
PN WO200177375-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 06-APR-2001; 2001WO-EP03968.  
XX  
PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-017470/02.  
XX  
PT New nucleic acid sequences from chemically modified genes associated  
PT with gene regulation, useful for analysing cytosine methylations for  
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency  
PT disease

XX  
PS Claim 1; SEQ ID No 165; 26pp; English.  
XX  
CC The invention relates to 224 nucleic acid sequences comprising at least  
CC 18 bases of a chemically pretreated gene associated with gene regulation  
CC selected from 43 known genes (or complementary sequences). The  
CC chemical pretreatment converts cytosine bases unmethylated at the  
CC 5-position to uracil or another base with hybridisation behaviour  
CC dissimilar to cytosine, to enable analysis of cytosine methylations.  
CC The DNA sequences, oligomers (or sets/arrays) and method are  
CC useful in the diagnosis of diseases (or predisposition to diseases)  
CC associated with gene regulation and in therapy of such diseases, by  
CC enabling analysis of the cytosine methylation patterns of such genes,  
CC kits are provided. They are especially useful in diagnosis  
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac  
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,  
CC preeclampsia, graft versus-host disease. The present sequence is a  
CC sequence included in the sequence data for this specification and is  
CC associated with the human gene regulation-associated genes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 5575 BP; 1212 A; 190 C; 1534 G; 2639 T; 0 other;  
XX  
Alignment Scores: 149 Length: 5575  
Pred. No.: 55.00 Matches: 8  
Score: 80.00% Conservative: 4  
Percent Similarity: 53.33% Mismatches: 3  
Best Local Similarity: 56.12% Indels: 0  
Query Match: 24 Gaps: 0  
DB: 1  
US-09-854-133-587 (1-16) x AAS61206 (1-5575)  
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePhe 15  
Db 1697 TTCGAAGGAAATGCGGGTTAGATGTTTGTATTTTATTTT 1741  
RESULT 9  
ABK84796/C  
ID ABK84796 standard; cDNA; 97835 BP.  
XX  
AC ABK84796;  
XX  
DT 14-AUG-2002 (first entry)  
XX  
DE Human cDNA differentially expressed in granulocytic cells #1367.  
XX  
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
XX  
OS Homo sapiens.  
XX  
PN WO200228999-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 03-OCT-2001; 2001WO-US30821.  
XX  
PR 03-OCT-2000; 2000US-237189P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;



XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

XX Claim 1; SEQ ID NO 10930; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3171 BP; 955 A; 806 C; 740 G; 670 T; 0 other;

#### Alignment Scores:

Pred. No.: 232 Length: 3171  
Score: 52.00 Matches: 8  
Percent Similarity: 62.50% Conservative: 2  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 53.06% Indels: 0  
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x ABL19819 (1-3171)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
Db 284 TTCCTCGCTGGACATGGATTGCTCTTTGTGTGCTTCTGCTCTTCTGG 237

#### RESULT 12

ABL16345/C

ID ABL16345 standard; DNA; 3430 BP.

XX ABL16345;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 508.

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 508.

XX Drosophila melanogaster.

XX WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

XX Claim 1; SEQ ID NO 508; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3430 BP; 1028 A; 859 C; 812 G; 731 T; 0 other;

#### Alignment Scores:

Pred. No.: 253 Length: 3430  
Score: 52.00 Matches: 8  
Percent Similarity: 62.50% Conservative: 2  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 53.06% Indels: 0  
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x ABL16345 (1-3430)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
Db 543 TTCCTCGCTGGACATGGATTGCTCTTTGTGTGCTTCTGCTCTTCTGG 496

#### RESULT 13

ABL32703

ID ABL32703 standard; DNA; 6089 BP.

XX ABL32703;

DT 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 676.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

PN 03-JAN-2002.

PD 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation -

XX Claim 1; SEQ ID NO 676; 32pp + Sequence Listing; German.

PS The present invention provides a number of human immune system associated

XX genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, anaemia, cancer, acute myeloid

CC macular degeneration, arteriosclerosis, AIDS, epilepsy, neurofibromatosis,

CC leukaemia, Alzheimer's disease, psoriasis and inflammatory/ulcerative bowel

CC diseases. The present sequence is a gene of the invention.

XX Sequence 6089 BP; 1370 A; 240 C; 1823 G; 2656 T; 0 other;

SQ

Alignment Scores:

Pred. No.:	474	Length:	6089
Score:	52.00	Matches:	7
Percent Similarity:	90.91%	Conservative:	3
Best Local Similarity:	63.64%	Mismatches:	1
Query Match:	53.06%	Indels:	0
DB:	24	Gaps:	0

US-09-854-133-587 (1-16) x ABL32703 (1-6089)

QY 5 CysGlyIleAspPheIleIlePheTrpIlePhe 15

Db 4865 TGTGGTATTGAGTATTGATTTTGGTATTT 4897

RESULT 14

ABL33497

ID ABL33497 standard; DNA; 6285 BP.

XX ABL33497;

AC

XX 26-MAR-2002 (first entry)

DT

XX Human immune system associated gene SEQ ID NO: 1470.

DE

XX Human; immune system disease; cytosine methylation; antiasthmatic;

XX antiarteriosclerotic; antianaemic; cytostatic; nootropic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;

KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

KW gene; ds.

XX

OS Homo sapiens.

XX

XX WO200200928-A2.

PN

XX 03-JAN-2002.

PD

XX 02-JUL-2001; 2001WO-EP07537.

PF

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

PR

XX (EPIG-) EPIGENOMICS AG.

PA

XX Olek A, Piepenbrock C, Berlin K;

PI

XX WPI; 2002-130909/17.

XX

XX Nucleic acid comprising fragment of chemically modified gene, useful

PT for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation -

XX

XX Claim 1; SEQ ID NO 1470; 32pp + Sequence Listing; German.

PS The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences

CC

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

CC diseases. The present sequence is a gene of the invention.

XX Sequence 6285 BP; 1346 A; 263 C; 1715 G; 2961 T; 0 other;

SQ

Alignment Scores:

Pred. No.:	490	Length:	6285
Score:	52.00	Matches:	9
Percent Similarity:	69.23%	Conservative:	0
Best Local Similarity:	69.23%	Mismatches:	4
Query Match:	53.06%	Indels:	0
DB:	24	Gaps:	0

US-09-854-133-587 (1-16) x ABL33497 (1-6285)

QY 4 AsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16

Db 4522 AATTAGGGAATAGATTATTTTGGGATTTGG 4560

RESULT 15

ABL19818

ID ABL19818 standard; DNA; 8164 BP.

XX ABL19818;

AC

XX 26-MAR-2002 (first entry)

DT

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 10927.

DE

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

KW

XX Drosophila melanogaster.

OS

XX WO200171042-A2.

PN

XX 27-SEP-2001.

PD

XX 23-MAR-2001; 2001WO-US09231.

PF

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PR

XX (PEKE ) PE CORP NY.

PA

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

PS Claim 1; SEQ ID NO 10927; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

XX Sequence 8164 BP; 2296 A; 1589 C; 1831 G; 2448 T; 0 other;

SQ



Alignment Scores:  
Pred. No.: 653 Length: 8164  
Score: 52.00 Matches: 8  
Percent Similarity: 62.50% Conservative: 2  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 53.06% Indels: 0  
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x ABL19818 (1-8164)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
||| ||| ||| |||:::|||||  
Db 6218 TTCCTCGCTGGACATGGATTGCTTTGTGTGCTTCGGCTCTCTGG 6265

Search completed: May 11, 2003, 14:16:32  
Job time : 49.6903 secs



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 14:25:24 ; Search time 8.77876 Seconds  
(without alignments)  
558.943 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 98  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues  
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_173956\_361/app\_query.fasta\_1.462  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn\_1\_1\_47 @runat\_05052003\_173956\_361 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match %	Length	ID	Description
C 1	59	60.2	17656	4 US-09-433-579-3	Sequence 3, Appli
C 2	53	54.1	40000	4 US-09-780-049-18	Sequence 18, Appli
C 3	48	49.0	4673	1 US-07-638-431-1	Sequence 1, Appli
C 4	48	49.0	4673	5 PCT-US92-00018-1	Sequence 1, Appli
C 5	46	46.9	66	3 US-09-046-247-138	Sequence 138, App
C 6	46	46.9	4370	4 US-08-981-527A-20	Sequence 20, Appli
C 7	45	45.9	885	3 US-08-545-196B-20	Sequence 20, Appli
C 8	45	45.9	1725	1 US-08-257-073-14	Sequence 14, Appli
C 9	45	45.9	2435	4 US-09-484-970B-134	Sequence 134, App
C 10	45	45.9	3719	1 US-08-920-812-10	Sequence 10, Appli
C 11	45	45.9	3719	1 US-08-920-827-10	Sequence 10, Appli
C 12	45	45.9	3719	1 US-08-921-177-10	Sequence 10, Appli

C 13	45	45.9	3719	1 US-08-362-577C-10	Sequence 10, Appli
C 14	45	45.9	3719	2 US-08-920-828-10	Sequence 10, Appli
C 15	45	45.9	4810	4 US-09-596-824-5	Sequence 5, Appli
C 16	45	45.9	6718	2 US-08-962-284-1	Sequence 1, Appli
C 17	44	44.9	840	5 PCT-US91-08177-12	Sequence 12, Appli
C 18	44	44.9	1001	4 US-09-641-638-461	Sequence 461, App
C 19	44	44.9	1725	4 US-09-134-001C-545	Sequence 545, App
C 20	44	44.9	1899	4 US-08-965-762-12	Sequence 12, Appli
C 21	44	44.9	6131	1 US-07-732-242C-8	Sequence 8, Appli
C 22	44	44.9	7152	4 US-09-167-681-29	Sequence 29, Appli
C 23	44	44.9	7323	5 PCT-US91-08177-1	Sequence 1, Appli
C 24	43	43.9	630	4 US-09-328-111-121	Sequence 121, App
C 25	43	43.9	876	4 US-08-446-137B-3	Sequence 3, Appli
C 26	43	43.9	2088	1 US-08-332-838-1	Sequence 1, Appli
C 27	43	43.9	2189	3 US-08-846-020A-1	Sequence 1, Appli
C 28	43	43.9	2189	4 US-09-617-871-1	Sequence 1, Appli
C 29	43	43.9	3279	4 US-08-446-137B-1	Sequence 1, Appli
C 30	43	43.9	3387	1 US-08-468-557-1	Sequence 1, Appli
C 31	43	43.9	11443	4 US-08-961-527-49	Sequence 49, Appli
C 32	43	43.9	62804	4 US-09-800-960-3	Sequence 3, Appli
C 33	42.5	43.4	338	4 US-09-615-192A-158	Sequence 158, App
C 34	42.5	43.4	1814	2 US-08-483-151-1	Sequence 1, Appli
C 35	42.5	43.4	1814	5 PCT-US96-06427-1	Sequence 1, Appli
C 36	42.5	43.4	5099	1 US-08-487-890A-4	Sequence 4, Appli
C 37	42.5	43.4	5099	2 US-08-478-435-4	Sequence 4, Appli
C 38	42.5	43.4	5099	2 US-08-337-483-4	Sequence 4, Appli
C 39	42.5	43.4	5099	2 US-08-478-373-4	Sequence 4, Appli
C 40	42.5	43.4	5099	3 US-08-474-671-4	Sequence 4, Appli
C 41	42.5	43.4	5099	3 US-08-483-577A-4	Sequence 4, Appli
C 42	42.5	43.4	5099	4 US-08-897-438-4	Sequence 4, Appli
C 43	42.5	43.4	5099	4 US-08-637-654-4	Sequence 4, Appli
C 44	42.5	43.4	5099	4 US-08-649-518-4	Sequence 4, Appli
C 45	42	42.9	830	4 US-08-998-416-419	Sequence 419, App

ALIGNMENTS

RESULT 1  
US-09-433-579-3/c  
; Sequence 3, Application US/09433579  
; Patent No. 6444877  
; GENERAL INFORMATION:  
; APPLICANT: Rottmann, William H.  
; TITLE OF INVENTION: LSAG Gene  
; FILE REFERENCE: LSAG Gene  
; CURRENT APPLICATION NUMBER: US/09/433,579  
; CURRENT FILING DATE: 1999-11-04  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 17656  
; TYPE: DNA  
; ORGANISM: Liquidambar styraciflua  
US-09-433-579-3

Alignment Scores:  
Pred. No.: 13.7 Length: 17656  
Score: 59.00 Matches: 11  
Percent Similarity: 44.00% Conservative: 0  
Best Local Similarity: 44.00% Mismatches: 2  
Query Match: 60.20% Indels: 12  
DB: 4 Gaps: 1

US-09-854-133-587 (1-16) x US-09-433-579-3 (1-17656)

QY 4 AsnCysGlyIle-----AspPheIleIle 11  
|||||  
Db 2548 AATTGTGGGATCCTCCATCACCAGATGCCAATCATACTTTTACTTTTCGATTTTATT 2489  
QY 12 PheTrpIlePheTrp 16  
|||||  
Db 2488 TTTTGGTTTTTTGG 2474

RESULT 2  
US-09-780-049-18  
; Sequence 18, Application US/09780049  
; Patent No. 6465250  
; GENERAL INFORMATION:  
; APPLICANT: Brett P. Monia  
; APPLICANT: Jacqueline Wyatt  
; TITLE OF INVENTION: ANTISENSE MODULATION OF PROTEIN PHOSPHATASE 2 CATALYTIC SUBUNIT  
; TITLE OF INVENTION: EXPRESSION  
; FILE REFERENCE: RTS-0134  
; CURRENT APPLICATION NUMBER: US/09/780,049  
; CURRENT FILING DATE: 2001-02-09  
; NUMBER OF SEQ ID NOS: 96  
; SEQ ID NO 18  
; LENGTH: 40000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-780-049-18

Alignment Scores:		
Pred. No.:	322	40000
Score:	53.00	7
Percent Similarity:	80.00%	5
Best Local Similarity:	46.67%	3
Query Match:	54.08%	0
DB:	4	0
		Gaps:
		Indels:
		Mismatches:
		Conservative:
		Matches:
		Length:

US-09-854-133-587 (1-16) x US-09-780-049-18 (1-40000)

**Oy**    2 GlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
       |||            ++++++ ::::||||:||||:||||:  
**Db** 16687 CAAAACCAAGTGGTACTAATACATTCTTTTCTTTTGTGG 16731

```

RESULT 3
US-07-638-431-1/c
; Sequence 1, Application US/07638431
; Patent No. 5198535
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khusmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: immunogen and gene
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMRDC Building 1 T-12 National Naval
; STREET: Medical Center
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/638,431
; FILING DATE: 19910110
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avrom D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4673 base pairs
; TYPE: NUCLEIC ACID

```

```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
; STRAIN: 17X(NL)
; DEVELOPMENTAL STAGE: erythrocytic stage
; TISSUE TYPE: Blood
; CELL TYPE: erythrocytic stage
; IMMEDIATE SOURCE:
; LIBRARY: Py-lambda gtl1-2-7 kb genomic expression
; CLONE: Py10.1111
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 718..3195
; OTHER INFORMATION:
;
; US-07-638-431-1

```

Alignment Scores:		
Pred. No.:	184	4673
Score:	48.00	6
Percent Similarity:	100.00%	2
Best Local Similarity:	75.00%	0
Query Match:	48.98%	0
DB:	1	0
	Length:	
	Matches:	
	Conservative:	
	Mismatches:	
	Indels:	
	Gaps:	

US-09-854-133-587 (1-16) x US-07-638-431-1 (1-4673)

Qy 9 PheIleIlePheTrpIlePheTrp 16  
|||  
Db 2601 TTCATTATCTCTGGGTTTATTGG 2578

RESULT 4  
PCT-US92-00018-1/c  
; Sequence 1, Application PC/TUS9200018  
; GENERAL INFORMATION:  
; APPLICANT: Hoffman, Stephen L.  
; APPLICANT: Charoenvit, Yupin  
; APPLICANT: Hedstrom, Richard  
; APPLICANT: Khumsmith, Srisin  
; APPLICANT: Rogers IV, William O.  
; TITLE OF INVENTION: Protective malaria sporozoite surface protein  
; TITLE OF INVENTION: immunogen and gene encoding  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: A. David Spevack  
; STREET: NMRDC Building 1 T-12 National Naval  
; STREET: Medical Center  
; CITY: Bethesda  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20814-5044  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/00018  
; FILING DATE: 19920103  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spevack, Avram D.  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 295-6759  
; TELEFAX: (301) 295-4033  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4673 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double



```

; FILING DATE: 8-SEPTEMBER-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/931,473
; FILING DATE: 17-AUGUST-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX 34.2/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA
; FEATURE:
; OTHER INFORMATION: All pyrimidines are 2'-F modified
US-09-046-247-138

Alignment Scores:
Pred. No.: 3.19 Length: 66
Score: 46.00 Matches: 7
Percent Similarity: 66.67% Conservative: 1
Best Local Similarity: 58.33% Mismatches: 4
Query Match: 46.94% Indels: 0
DB: 3 Gaps: 0

US-09-854-133-587 (1-16) x US-09-046-247-138 (1-66)

Qy 5 CysGlyIleAspPheIleIlePheTrpIlePheTrp 16
      |||||
Db 11 UGCGUAUUGACUUUUGUUUUUUUUUUUGCCUGG 46

RESULT 6
US-08-981-527A-20
; Sequence 20, Application US/08981527A
; Patent No. 6410262
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Kerkman, Richard
; APPLICANT: Broekhuizen, Cornelis P.
; TITLE OF INVENTION: No. 6410262el Secretion Factors for
; TITLE OF INVENTION: Gram-Positive Microorganisms Genes E
; TITLE OF INVENTION: of Using It
; FILE REFERENCE: GCX322-US
; CURRENT APPLICATION NUMBER: US/08/981,527A
; CURRENT FILING DATE: 1997-12-16
; PRIOR APPLICATION NUMBER: PCT/NL96/00278
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 4370
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-08-981-527A-20

Alignment Scores:
Pred. No.: 360 Length: 4370
Score: 46.00 Matches: 8
Percent Similarity: 75.00% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 4
Query Match: 46.94% Indels: 0
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-08-981-527A-20 (1-4370)

Qy 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp
      |||
Db 3839 TTCATATACAAATATCGGAATAAATGGATGATATTAGCGTATTTGG

```

```
RESULT 7
US-08-545-196B-20/c
; Sequence 20, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 18..881
US-08-545-196B-20
Alignment Scores:
Pred. No.: 86.4 Length: 885
Score: 45.00 Matches: 10
Percent Similarity: 46.43% Conservative: 3
Best Local Similarity: 35.71% Mismatches: 3
Query Match: 45.92% Indels: 12
DB: 3 Gaps: 2
US-09-854-133-587 (1-16) x US-08-545-196B-20 (1-885)
QY 1 PheGlnAlaAsnCysGlyIle-----AspPheIleIle----- 11
Db 278 TTCAAGGGAGTTGTGGCATCTCTTTTGGCTTTTATCTCTTGGCAGGTTTCTTCT 219
QY 12 -----PheTrpIlePheTrp 16
Db 218 GGCTGTGCCTTTTGGCTATCTGG 195
RESULT 8
US-08-257-073-14/c
; Sequence 14, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paolletti, Enzo
; APPLICANT: de Taisné, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
```

```
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1725 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-257-073-14
Alignment Scores:
Pred. No.: 183 Length: 1725
Score: 45.00 Matches: 6
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 85.71% Mismatches: 0
Query Match: 45.92% Indels: 0
DB: 1 Gaps: 0
US-09-854-133-587 (1-16) x US-08-257-073-14 (1-1725)
QY 10 IleIlePheTrpIlePheTrp 16
Db 1194 ATCGTCTCTGGATTTTGG 1174
RESULT 9
US-09-484-970B-134
; Sequence 134, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 134
; LENGTH: 2435
; TYPE: DNA
```

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
OTHER INFORMATION: Incyte ID No. 6426186 247789.2CBI  
NAME/KEY: unsure  
LOCATION: 93, 128, 132, 143-144, 2419, 2427, 2429  
OTHER INFORMATION: a, t, c, g, or other  
US-09-484-970B-134

Alignment Scores:  
Pred. No.: 270 Length: 2435  
Score: 45.00 Matches: 6  
Percent Similarity: 90.00% Conservative: 3  
Best Local Similarity: 60.00% Mismatches: 1  
Query Match: 45.92% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-484-970B-134 (1-2435)

QY 6 GlyIleAspPheIleIlePheTrpIlePhe 15

Db 1253 GGGATCGACTGGTTCCTCTTCTGGGTCTTC 1282

RESULT 10

US-08-920-812-10/c  
Sequence 10, Application US/08920812  
Patent No. 5763188

GENERAL INFORMATION:

APPLICANT: Ohno, Tsuneya  
APPLICANT: Matsuhisa, Akio  
APPLICANT: Uehara, Hirotsugu  
APPLICANT: Ega, Soji

TITLE OF INVENTION: Probe for Diagnosing Infectious Disease

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/920,812

FILING DATE: 29-AUG-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/362,577

FILING DATE: 27-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien

REGISTRATION NUMBER: 33,547

REFERENCE/DOCKET NUMBER: 19036/32420

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 3719 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecalis

STRAIN: Clinical Isolate S2-3

US-08-920-812-10

Alignment Scores:  
Pred. No.: 435 Length: 3719  
Score: 45.00 Matches: 7  
Percent Similarity: 54.55% Conservative: 5  
Best Local Similarity: 31.82% Mismatches: 2  
Query Match: 45.92% Indels: 8  
DB: 1 Gaps: 1

US-09-854-133-587 (1-16) x US-08-920-812-10 (1-3719)

QY 1 PheGlnAlaAsnCysGlyIle-----AspPheIleIlePhe 12

Db 1878 TTTAGAACGGACTGTGGTGTGTTGTTTGACGAACACCGGAAGCGACTTCTTTTGT 1819

QY 13 TrpIle 14

Db 1818 TGGCTT 1813

RESULT 11

US-08-920-827-10/c

Sequence 10, Application US/08920827

Patent No. 5770375

GENERAL INFORMATION:

APPLICANT: Ohno, Tsuneya

APPLICANT: Matsuhisa, Akio

APPLICANT: Uehara, Hirotsugu

APPLICANT: Ega, Soji

TITLE OF INVENTION: Probe for Diagnosing Infectious Disease

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/920,827

FILING DATE: 29-AUG-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/362,577

FILING DATE: 27-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien

REGISTRATION NUMBER: 33,547

REFERENCE/DOCKET NUMBER: 19036/32420

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 3719 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecalis

STRAIN: Clinical Isolate S2-3

US-08-920-827-10

Alignment Scores:

Pred. No.: 435 Length: 3719

Score: 45.00 Matches: 7

Percent Similarity: 54.55% Conservative: 5

Best Local Similarity: 31.82% Mismatches: 2

Query Match: 45.92% Indels: 8  
DB: 1 Gaps: 1  
US-09-854-133-587 (1-16) x US-08-920-827-10 (1-3719)  
QY 1 PheGlnAlaAsnCysGlyIle-----AspPheIleIlePhe 12  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | |  
Db 1878 TTTAGAACGGACTGTGGTGTGTTTGTGACGAACCTACCGGAAGCGACTTCTTTTGTGTTT 1819  
QY 13 TrpIle 14  
| | | | |  
Db 1818 TGGCTT 1813  
RESULT 12  
US-08-921-177-10/c  
; Sequence 10, Application US/08921177  
; Patent No. 5798211  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotsugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/921,177  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3719 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORGANISM: Enterococcus faecalis  
; STRAIN: Clinical Isolate S2-3  
US-08-921-177-10  
Alignment Scores:  
Pred. No.: 435 Length: 3719  
Score: 45.00 Matches: 7  
Percent Similarity: 54.55% Conservative: 5  
Best Local Similarity: 31.82% Mismatches: 2  
Query Match: 45.92% Indels: 8  
DB: 1 Gaps: 1  
US-09-854-133-587 (1-16) x US-08-921-177-10 (1-3719)

QY 1 PheGlnAlaAsnCysGlyIle-----AspPheIleIlePhe 12  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | |  
Db 1878 TTTAGAACGGACTGTGGTGTGTTTGTGACGAACCTACCGGAAGCGACTTCTTTTGTGTTT 1819  
QY 13 TrpIle 14  
| | | | |  
Db 1818 TGGCTT 1813  
RESULT 13  
US-08-362-577C-10/c  
; Sequence 10, Application US/08362577C  
; Patent No. 5807673  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotsugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,577C  
; FILING DATE: 27-MAR-1995  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3719 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecalis  
; STRAIN: Clinical Isolate S2-3  
US-08-362-577C-10  
Alignment Scores:  
Pred. No.: 435 Length: 3719  
Score: 45.00 Matches: 7  
Percent Similarity: 54.55% Conservative: 5  
Best Local Similarity: 31.82% Mismatches: 2  
Query Match: 45.92% Indels: 8  
DB: 1 Gaps: 1  
US-09-854-133-587 (1-16) x US-08-362-577C-10 (1-3719)  
QY 1 PheGlnAlaAsnCysGlyIle-----AspPheIleIlePhe 12  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | |  
Db 1878 TTTAGAACGGACTGTGGTGTGTTTGTGACGAACCTACCGGAAGCGACTTCTTTTGTGTTT 1819  
QY 13 TrpIle 14  
| | | | |  
Db 1818 TGGCTT 1813



```

RESULT 14
US-08-920-828-10/c
; Sequence 10, Application US/08920828
; Patent No. 5853998
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,828
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3719 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecalis
; STRAIN: Clinical Isolate S2-3
US-08-920-828-10

Alignment Scores:
Pred. No.: 435 Length: 3719
Score: 45.00 Matches: 7
Percent Similarity: 54.55% Conservative: 5
Best Local Similarity: 31.82% Mismatches: 2
Query Match: 45.92% Indels: 8
DB: 2 Gaps: 1

US-09-854-133-587 (1-16) x US-08-920-828-10 (1-3719)
QY 1 PheGlnAlaAsnCysGlyIle-----AspPheIleIlePhe 12
||||: |||:||||:
Db 1878 TTTAGACGGACTGTGGTGTGTTTTCACGAACACTACCGGAAGCGACTCTTTTGTGTTT 1819
QY 13 TrpIle 14
||||:
Db 1818 TGGCTT 1813

RESULT 15
US-09-596-824-5/c
; Sequence 5, Application US/09596824
; Patent No. 6372464
; GENERAL INFORMATION:

```

```

; APPLICANT: Debbie S. Yaver
; APPLICANT: Barbara McArdle
; TITLE OF INVENTION: Polypeptides Having Peroxidase Activity
; TITLE OF INVENTION: And Nucleic Acids Encoding Same
; FILE REFERENCE: 10048.000-US
; CURRENT APPLICATION NUMBER: US/09/596,824
; CURRENT FILING DATE: 2000-06-19
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 4810
; TYPE: DNA
; ORGANISM: Bjerkandera adusta
US-09-596-824-5

Alignment Scores:
Pred. No.: 582 Length: 4810
Score: 45.00 Matches: 5
Percent Similarity: 69.23% Conservative: 4
Best Local Similarity: 38.46% Mismatches: 4
Query Match: 45.92% Indels: 0
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-596-824-5 (1-4810)
QY 4 AsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
::: ||||| |||:|:|: |||:|:|:
Db 1180 TCGCGGGGCATTTCCTTTGTTGTGGCATGGCTGATATGG 1142

Search completed: May 11, 2003, 16:16:10
Job time : 15.7788 secs

```



GenCore version 5.1.4\_p5\_4578  
 Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 15:28:15 ; Search time 18.9735 Seconds  
 (without alignments)  
 1047.953 Million cell updates/sec

Title: US-09-854-133-587  
 Perfect score: 98  
 Sequence: 1 FOANCGIDFIIFWIFW 16

Scoring table: BLOSUM62  
 Xgapop 10.0 , Xgapext 0.5  
 Ygapop 10.0 , Ygapext 0.5  
 Fgapop 6.0 , Fgapext 7.0  
 Delop 6.0 , Delext 7.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Command line parameters:  
 -MODEL=frame+p2n.model -DEV=xlh  
 -Q=/cgn2\_1/USPTO.spool/US09854133/runat\_05052003\_173958\_449/app\_query.fasta\_1.462  
 -DB=PublishedApplications\_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosu62  
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0  
 -MAXLEN=2000000000 -USER=US09854133@cgn\_1\_1\_117@runat\_05052003\_173958\_449  
 -NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
 -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications\_NA:

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	98	100.0	337	9	US-09-854-133-442
2	98	100.0	337	10	US-09-738-973-442
3	98	100.0	2239	9	US-09-854-133-440
4	98	100.0	2239	10	US-09-738-973-440

5	98	100.0	5981	9	US-09-854-133-441	Sequence 441, App
6	98	100.0	5981	10	US-09-738-973-441	Sequence 441, App
7	51	52.0	806	10	US-09-910-943-590	Sequence 590, App
8	51	52.0	31412	9	US-10-109-551-3	Sequence 3, Appli
9	50.5	51.5	710	10	US-09-867-550-1537	Sequence 1537, Ap
10	50	51.0	8925	9	US-09-764-891-9932	Sequence 9932, Ap
11	50	51.0	15732	9	US-10-239-676-96	Sequence 96, Appl
12	50	51.0	1691139	9	US-10-067-514-1	Sequence 1, Appli
13	49	50.0	2157	10	US-09-801-368-387	Sequence 387, App
14	49	50.0	83450	9	US-09-811-469-3	Sequence 3, Appli
15	48	49.0	303	10	US-09-998-598-2471	Sequence 2471, Ap
16	48	49.0	405	9	US-09-918-995-33363	Sequence 33363, A
17	48	49.0	601	10	US-09-777-921A-64	Sequence 64, Appl
18	48	49.0	2285	9	US-09-978-295A-283	Sequence 283, App
19	48	49.0	2285	9	US-09-978-697-283	Sequence 283, App
20	48	49.0	2285	9	US-09-978-192A-283	Sequence 283, App
21	48	49.0	2285	9	US-09-999-832A-283	Sequence 283, App
22	48	49.0	2285	9	US-09-978-189-283	Sequence 283, App
23	48	49.0	2285	9	US-10-174-590-111	Sequence 111, App
24	48	49.0	2285	9	US-10-176-758-111	Sequence 111, App
25	48	49.0	2285	9	US-10-175-737-111	Sequence 111, App
26	48	49.0	2285	9	US-10-173-706-111	Sequence 111, App
27	48	49.0	2285	9	US-10-175-738-111	Sequence 111, App
28	48	49.0	2285	9	US-10-175-752-111	Sequence 111, App
29	48	49.0	2285	9	US-10-176-482-111	Sequence 111, App
30	48	49.0	2285	9	US-10-176-757-111	Sequence 111, App
31	48	49.0	2285	9	US-10-176-913-111	Sequence 111, App
32	48	49.0	2285	9	US-10-180-552-111	Sequence 111, App
33	48	49.0	2285	9	US-10-180-557-111	Sequence 111, App
34	48	49.0	2285	9	US-10-173-700-111	Sequence 111, App
35	48	49.0	2285	9	US-10-174-572-111	Sequence 111, App
36	48	49.0	2285	9	US-10-174-579-111	Sequence 111, App
37	48	49.0	2285	9	US-10-174-582-111	Sequence 111, App
38	48	49.0	2285	9	US-10-174-588-111	Sequence 111, App
39	48	49.0	2285	9	US-10-175-739-111	Sequence 111, App
40	48	49.0	2285	9	US-10-175-740-111	Sequence 111, App
41	48	49.0	2285	9	US-10-175-743-111	Sequence 111, App
42	48	49.0	2285	9	US-10-176-488-111	Sequence 111, App
43	48	49.0	2285	9	US-10-176-492-111	Sequence 111, App
44	48	49.0	2285	9	US-10-176-747-111	Sequence 111, App
45	48	49.0	2285	9	US-10-176-750-111	Sequence 111, App

ALIGNMENTS

RESULT 1

US-09-854-133-442

; Sequence 442, Application US/09854133

; Publication No. US20020183499A1

; GENERAL INFORMATION:

; APPLICANT: Lodes, Michael J.

; APPLICANT: Mohamath, Raodoh

; APPLICANT: Henderson, Robert A.

; APPLICANT: Benson, Darin R.

; APPLICANT: Secrist, Heather

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.475C10

; CURRENT APPLICATION NUMBER: US/09/854,133

; CURRENT FILING DATE: 2001-05-11

; NUMBER OF SEQ ID NOS: 735

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 442

; LENGTH: 337

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-854-133-442

Alignment Scores:

Pred. No.: 2.04e-07 Length: 337

Score: 98.00 Matches: 16

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

```
; ORGANISM: Homo sapiens
US-09-854-133-440

Alignment Scores:
Pred. No.: 1.91e-06      Length: 2239
Score: 98.00           Matches: 16
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 9                  Gaps: 0

US-09-854-133-587 (1-16) x US-09-854-133-442 (1-337)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
    |||||||
Db 107 TTCCAGGCCAATTGTGGCATAGATTTTATCATATTTCTGGATTTTGG 154

RESULT 2
US-09-738-973-442
; Sequence 442, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 442
; LENGTH: 337
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-738-973-442

Alignment Scores:
Pred. No.: 2.04e-07      Length: 337
Score: 98.00           Matches: 16
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 10                  Gaps: 0

US-09-854-133-587 (1-16) x US-09-738-973-442 (1-337)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
    |||||||
Db 107 TTCCAGGCCAATTGTGGCATAGATTTTATCATATTTCTGGATTTTGG 154

RESULT 3
US-09-854-133-440
; Sequence 440, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 440
; LENGTH: 2239
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-09-854-133-440

Alignment Scores:
Pred. No.: 1.91e-06      Length: 2239
Score: 98.00           Matches: 16
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 9                  Gaps: 0

US-09-854-133-587 (1-16) x US-09-854-133-440 (1-2239)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
    |||||||
Db 104 TTCCAGGCCAATTGTGGCATAGATTTTATCATATTTCTGGATTTTGG 151

RESULT 4
US-09-738-973-440
; Sequence 440, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 440
; LENGTH: 2239
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-738-973-440

Alignment Scores:
Pred. No.: 1.91e-06      Length: 2239
Score: 98.00           Matches: 16
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 10                  Gaps: 0

US-09-854-133-587 (1-16) x US-09-738-973-440 (1-2239)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
    |||||||
Db 104 TTCCAGGCCAATTGTGGCATAGATTTTATCATATTTCTGGATTTTGG 151

RESULT 5
US-09-854-133-441
; Sequence 441, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
```



Patent No. US20020081610A1

```
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1537
; LENGTH: 710
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-867-550-1537

Alignment Scores:
Pred. No.: 20.6 Length: 710
Score: 50.50 Matches: 8
Percent Similarity: 71.43% Conservative: 2
Best Local Similarity: 57.14% Mismatches: 3
Query Match: 51.53% Indels: 1
DB: 10 Gaps: 1

US-09-854-133-587 (1-16) x US-09-867-550-1537 (1-710)

QY 4 AsnCysGly---IleAspPheIleIlePheTrpIlePheTrp 16
   ||| ||| :: |||||::||| |||||
Db 174 AACGCTGCAGGGTTAGTTTCATTTGTTTGGTTTGG 215

RESULT 10
US-09-764-891-9932/c
; Sequence 9932, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 9932
; LENGTH: 8925
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9932

Alignment Scores:
Pred. No.: 492 Length: 8925
Score: 50.00 Matches: 8
Percent Similarity: 71.43% Conservative: 2
Best Local Similarity: 57.14% Mismatches: 4
Query Match: 51.02% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-764-891-9932 (1-8925)

QY 3 AlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
   |||:: ||||| |||||::|||
Db 5144 GCATCATCTGCGATAGACAGATACATCTTCTGGCTATTCTGG 5103

RESULT 11
US-10-239-676-96
; Sequence 96, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 96
; LENGTH: 15732
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-96

Alignment Scores:
Pred. No.: 962 Length: 15732
Score: 50.00 Matches: 7
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 58.33% Mismatches: 3
Query Match: 51.02% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-10-239-676-96 (1-15732)

QY 4 AsnCysGlyIleAspPheIleIlePheTrpIlePhe 15
   ::|||::||| ||||| ||||| |||
Db 475 AGTTGTGGAATGGATTAGTTTGGTTTGGTTT 510

RESULT 12
US-10-067-514-1
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
; APPLICANT: Gretarsdottir, Solveig
; APPLICANT: Jonsdottir, Sif
; APPLICANT: Reynisdottir, Sigrídur Th.
; TITLE OF INVENTION: HUMAN STROKE GENE
; FILE REFERENCE: 2345.2010-003
; CURRENT APPLICATION NUMBER: US/10/067,514
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 09/811/352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Human
US-10-067-514-1

Alignment Scores:
Pred. No.: 2.24e+05 Length: 1691139
Score: 50.00 Matches: 7
Percent Similarity: 73.33% Conservative: 4
Best Local Similarity: 46.67% Mismatches: 2
Query Match: 51.02% Indels: 2
DB: 9 Gaps: 1

US-09-854-133-587 (1-16) x US-10-067-514-1 (1-1691139)

QY 4 AsnCysGlyIle-----AspPheIleIlePheTrpIlePheTrp 16
```







GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 14:07:50 ; Search time 242.549 seconds  
(without alignments)  
1068.354 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 98  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09854133/runat\_05052003\_173956\_349/app\_query.fasta\_1.462  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn\_1\_1\_2013\_erunat\_05052003\_173956\_349 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
C 1	59	60.2	813	17	BH437997	BH437997 BOHJI44TR
C 2	57	58.2	322	9	AV169412	AV169412 AV169412
C 3	57	58.2	402	10	AW845948	AW845948 PM4-CT015
C 4	57	58.2	402	10	AW845949	AW845949 PM4-CT015
C 5	57	58.2	650	17	AZ340196	AZ340196 1M0072H24
C 6	57	58.2	751	12	BF274788	BF274788 GA_Eb002
C 7	57	58.2	1231	13	BI085630	BI085630 602870174
C 8	56	57.1	173	12	BF471824	BF471824 UI-M-BH3-
C 9	56	57.1	239	12	BE948302	BE948302 UI-M-BH3-
C 10	56	57.1	327	12	BF225111	BF225111 uz43b01.y
C 11	56	57.1	337	12	BF320777	BF320777 uz56b08.y
C 12	56	57.1	383	12	BE986193	BE986193 UI-M-CG0P
C 13	56	57.1	467	9	AA967175	AA967175 ua50c04.r
C 14	56	57.1	470	10	BE200533	BE200533 ug70c08.y
C 15	56	57.1	516	9	AI527491	AI527491 uf12c07.y
C 16	56	57.1	579	14	BQ564930	BQ564930 gi27e11.y
C 17	56	57.1	583	14	BQ086748	BQ086748 ih89cl1.y
C 18	56	57.1	588	12	BF136680	BF136680 601780445
C 19	56	57.1	599	10	AW213068	AW213068 un98h09.y
C 20	56	57.1	622	14	BQ921841	BQ921841 AGENCOURT
C 21	56	57.1	642	10	AW989255	AW989255 uf24d09.y
C 22	56	57.1	667	12	BF142738	BF142738 601789992
C 23	56	57.1	726	13	BG919850	BG919850 602823591
C 24	56	57.1	732	10	AW213203	AW213203 uo01d05.y
C 25	56	57.1	898	14	BQ878072	BQ878072 AGENCOURT
C 26	55	56.1	168	14	BQ375999	BQ375999 CM0-TN003
C 27	55	56.1	214	17	AQ005390	AQ005390 CIT-HSP-2
C 28	55	56.1	358	9	AI970001	AI970001 wq79c06.x
C 29	55	56.1	512	17	AQ768884	AQ768884 HS_3191_A
C 30	55	56.1	515	10	AW510674	AW510674 hc89d06.x
C 31	54	55.1	212	10	BE057036	BE057036 GI26 Giga
C 32	54	55.1	252	9	AV255114	AV255114 AV255114
C 33	54	55.1	583	10	AV719226	AV719226 AV719226
C 34	54	55.1	617	17	AZ004528	AZ004528 RPCI-23-3
C 35	54	55.1	820	17	BH063329	BH063329 RPCI-24-3
C 36	54	55.1	1062	17	CNS051TA	AL317287 Tetraodon
C 37	53	54.1	303	17	TA212D11Q	AI479576 T. brucei
C 38	53	54.1	481	17	BH277396	BH277396 CH230-108
C 39	53	54.1	807	12	BG839567	BG839567 Gm01_13h1
C 40	53	54.1	830	13	BI889060	BI889060 ZF637-2-0
C 41	53	54.1	972	17	AG133150	AG133150 Pan trogl
C 42	53	54.1	1012	17	AG138611	AG138611 Pan trogl
C 43	52.5	53.6	257	12	BF835845	BF835845 CM0-HT095
C 44	52.5	53.6	325	10	AV333550	AV333550 AV333550
C 45	52.5	53.6	688	9	AI067874	AI067874 EST209558

ALIGNMENTS

RESULT 1  
BH437997/C  
LOCUS BOHJI44TR BOHJ Brassica oleracea genomic clone BOHJI44, DNA  
DEFINITION BH437997 813 bp DNA linear GSS 12-DEC-2001  
sequence.  
ACCESSION BH437997  
VERSION BH437997.1 GI:17623718  
KEYWORDS GSS.  
SOURCE Brassica oleracea.  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 813)  
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNAL Unpublished (2001)  
COMMENT Other\_GSSs: BOHJI44TF  
Contact: Chris Town  
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.

FEATURES  
source

Location/Qualifiers

1. .813  
/organism="Brassica oleracea"  
/strain="T01000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHJI44"  
/note="Vector: phos1; Site\_1: BstXI; 2-3 kb sheared  
genomic DNA inserted into phos1 using BstXI linkers"

BASE COUNT 273 a 209 c 150 g 181 t  
ORIGIN

Alignment Scores:  
pred. No.: 102 Length: 813  
Score: 59.00 Matches: 8  
Percent Similarity: 75.00% Conservatives: 4  
Best Local Similarity: 50.00% Mismatches: 4  
Query Match: 60.20% Indels: 0  
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x BH437997 (1-813)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIlePheTrpIlePheTrp 16  
Db 708 TTTATGCCATGCTAGGCTAGATTCAATTATTATTTGGTTTCTGG 661

RESULT 2  
AV169412/c  
LOCUS AV169412 322 bp mRNA linear EST 06-JUL-1999  
DEFINITION AV169412 Mus musculus head C57BL/6J 13-day embryo Mus musculus cDNA  
clone 3110084I03, mRNA sequence.  
ACCESSION AV169412  
VERSION AV169412.1 GI:5375849  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 322)  
Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,  
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara  
A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,  
Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C.,  
Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara  
Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N.,  
Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,  
Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.

RIKEN Mouse ESTs  
Unpublished (1999)  
Contact: Chie Owa  
Genome Science Laboratory  
RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-9145  
Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp  
Thermotabilization and thermoactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length cDNA  
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

FEATURES  
source

Location/Qualifiers

1. .322  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="3110084I03"  
/clone\_lib="Mus musculus head C57BL/6J 13-day embryo"  
/sex="mixed"  
/tissue\_type="head"  
/dev\_stage="13-day embryo"  
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia  
with a modified polylinker; Site\_1: Not I; Site\_2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer [5',  
TGTTACCAATCTGAAGTGGGAGCGCGCGGAATGGTTTGTGTGTGTGTGTGTGT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."

BASE COUNT 110 a 81 c 31 g 100 t  
ORIGIN

Alignment Scores:  
pred. No.: 81.6 Length: 322  
Score: 57.00 Matches: 7  
Percent Similarity: 77.78% Conservatives: 7  
Best Local Similarity: 38.89% Mismatches: 0  
Query Match: 58.16% Indels: 4  
DB: 9 Gaps: 1

US-09-854-133-587 (1-16) x AV169412 (1-322)

QY 3 AlaAsnCysGly-----IleaspPheIlePheTrpIlePheTrp 16  
Db 157 AGCTCTTGGGTCGGCTGCTGTTAGTTGAATTTTGGTTTGGTTATTTGG 104

RESULT 3  
AW845948/c  
LOCUS AW845948 402 bp mRNA linear EST 19-MAY-2000  
DEFINITION PM4-CT0152-220999-001-A05 CT0152 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW845948  
VERSION AW845948.1 GI:7941465  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 402)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=PM4-CT0152-220

999-001-A05&t3=1999-09-22&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 402.  
Location/Qualifiers

FEATURES  
source

1. .402  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0152"  
/dev\_stage="Adult"  
/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 ,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 131 a 90 c 82 g 99 t  
ORIGIN

Alignment Scores: 100 Length: 402  
Pred. No.: 57.00 Matches: 8  
Score: 76.92%  
Percent Similarity: 76.92% Conservative: 2  
Best Local Similarity: 61.54% Mismatches: 3  
Query Match: 58.16% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x AW845948 (1-402)

QY 4 AsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
:::|||||:::|||||  
Db 212 TCCTGCGGAATTAACCTAGCCGATTTGGATATTGG 174

RESULT 4

AW845949/c

LOCUS

DEFINITION PM4-CT0152-220999-001-A08 CT0152 Homo sapiens cDNA, mRNA sequence. EST 19-MAY-2000

ACCESSION AW845949

VERSION AW845949.1 GI:7941466

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 402)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

{http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=PM4-CT0152-220

999-001-A08&t3=1999-09-22&t4=1}

Seq primer: puc 18 forward

High quality sequence stop: 402.

Location/Qualifiers

1. .402

/organism="Homo sapiens"

/db\_xref="taxon:9606"

FEATURES  
source

/clone\_lib="CT0152"  
/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 ,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 131 a 90 c 82 g 99 t  
ORIGIN

Alignment Scores:

Pred. No.: 100 Length: 402  
Score: 57.00 Matches: 8  
Percent Similarity: 76.92% Conservative: 2  
Best Local Similarity: 61.54% Mismatches: 3  
Query Match: 58.16% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x AW845949 (1-402)

QY 4 AsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
:::|||||:::|||||

Db 212 TCCTGCGGAATTAACCTAGCCGATTTGGATATTGG 174

RESULT 5

AZ340196/c

LOCUS

DEFINITION 1M0072H24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0072H24 F, DNA sequence. 650 bp linear GSS 29-SEP-2000

ACCESSION AZ340196

VERSION AZ340196.1 GI:10415403

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 650)

REFERENCE

AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0072 row: H column: 24

Seq primer: CGTTGTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 650.

FEATURES

source

1. .650

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0072H24"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: pWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a







was used as a driver in a hybridization with a pool of the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH\_BMAP\_M\_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG\_SEQ=None found"

BASE COUNT 86 a 63 c 43 g 47 t

ORIGIN

Alignment Scores:

Pred. No.: 85.1 Length: 239  
Score: 56.00 Matches: 7  
Percent Similarity: 66.67% Conservative: 3  
Best Local Similarity: 46.67% Mismatches: 5  
Query Match: 57.14% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BE948302 (1-239)

QY 2 GlnAlaAsnCysGlylleAspPheIlePheTrpIlePheTrp 16  
Db 134 CAGCTGACATGTGCAAAATTTTGTGTTGTTTGGGTTGG 90

RESULT 10  
BF225111  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BF225111  
uz43b01.y1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone linear EST 29-DEC-2000  
mRNA sequence.  
BF225111.1 GI:11132576  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 327)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Other\_ESTs: uz43b01.x1  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[image.llnl.gov/image/html/iresources.shtml](http://image.llnl.gov/image/html/iresources.shtml)

JOURNAL  
COMMENT  
MG1:1432553  
High quality sequence stop: 323.  
Location/Qualifiers  
1. .327  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3671785"  
/clone\_lib="NCI\_CGAP\_Mam6"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: Sali;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Jeffrey Green, M.D., NIH"

FEATURES

source

BASE COUNT 70 a 58 c 73 g 126 t

ORIGIN

Alignment Scores:

Pred. No.: 114 Length: 327  
Score: 56.00 Matches: 7  
Percent Similarity: 66.67% Conservative: 3  
Best Local Similarity: 46.67% Mismatches: 5  
Query Match: 57.14% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BF225111 (1-327)

QY 2 GlnAlaAsnCysGlylleAspPheIlePheTrpIlePheTrp 16  
Db 136 CAGCTGACATGTGCAAAATTTTGTGTTGTTTGGGTTGG 180

RESULT 11  
BF320777  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BF320777  
uz56b08.y1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone linear EST 29-DEC-2000  
mRNA sequence.  
BF320777.1 GI:11269789  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 337)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[image.llnl.gov/image/html/iresources.shtml](http://image.llnl.gov/image/html/iresources.shtml)

MG1:1433815  
High quality sequence stop: 329.  
Location/Qualifiers  
1. .337  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3673047"  
/clone\_lib="NCI\_CGAP\_Mam6"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: Sali;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Jeffrey Green, M.D., NIH"

FEATURES

source

BASE COUNT 84 a 53 c 89 g 111 t

ORIGIN

Alignment Scores:

Pred. No.: 117 Length: 337  
Score: 56.00 Matches: 7  
Percent Similarity: 66.67% Conservative: 3  
Best Local Similarity: 46.67% Mismatches: 5  
Query Match: 57.14% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BF320777 (1-337)

QY 2 GlnAlaAsnCysGlyIleAaspPheIleIlePheTrpIlePheTrp 16  
 ||| ||| |||:|||||:|||||  
 Db 289 CAGCTGACATGTGCAAAATTTTGTGTTGTTTTGGGTGTTGG 333

RESULT 12  
 BE986193/c

LOCUS  
 DEFINITION  
 BE986193  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

BE986193 383 bp mRNA linear EST 29-APR-2002  
 UI-M-CG0p-bgp-b-09-0-UI.s1 NIH\_BMAP\_Ret4\_S2 Mus musculus cDNA clone  
 UI-M-CG0p-bgp-b-09-0-UI 3', mRNA sequence.  
 BE986193  
 BE986193.1 GI:10660263  
 EST.  
 house mouse.  
 Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 383)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 97044477

Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mEst@mail.nih.gov

The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to identify it as a clone from the  
 retina tissue cDNA Library Preparation: M.B. Soares Lab Clone  
 distribution: Researchers may obtain BMAP cDNA clones from RESEARCH  
 GENETICS. It should be noted that Bento Soares is generating a  
 small number of additional specialized non-redundant arrays of BMAP  
 cDNAs whose availability will be considered under appropriate and  
 limited collaborative arrangements The tissue for this library was  
 contributed by Dr. Xin-Yuan Fu, Yale University School of Medicine  
 Seq primer: M13 Forward  
 POLYA=Yes.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT

BASE COUNT	127 a	84 c	65 g	105 t	2 others
ORIGIN					
Alignment Scores:					
Pred. No.:	132			Length:	383
Score:	56.00			Matches:	7
Percent Similarity:	66.67%			Conservative:	3
Best Local Similarity:	46.67%			Mismatches:	5
Query Match:	57.14%			Indels:	0

DB:	12	Gaps:	0
US-09-854-133-587 (1-16) x BE986193 (1-383)			
QY	2	GlnAlaAsnCysGlyIleAspPheIlePheTrpIlePheTrp 16	
Db	167	CAGCTGACATGTGCAAAATTTTGTGTGTTTGGGTGGTGG 123	
RESULT 13			
AA967175			
LOCUS	AA967175	467 bp	mRNA linear EST 19-MAY-1998
DEFINITION	ua50c04.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone		
	IMAGE:1350150 5' similar to gb:X65627 M.musculus mRNA TNZ2 for p68		
	RNA helicase (MOUSE);, mRNA sequence.		
ACCESSION	AA967175		
VERSION	AA967175.1	GI:3141068	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 467)		
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE	The WashU-HHMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:698942		
Seq primer:	-28ml3 rev1 ET from Amersham		
High quality sequence stop:	385.		
FEATURES	Location/Qualifiers		
source	1. .467		
	/organism="Mus musculus"		
	/strain="C57BL/6 x CBA"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:1350150"		
	/clone_lib="Stratagene mouse lung 937302"		
	/sex="female"		
	/tissue_type="lung"		
	/dev_stage="6-8 month old"		
	/lab_host="SOLR (kanamycin resistant)"		
	/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI		
	; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo		
	dT. 6-8 month old female lung and 1.5 year old male lung		
	were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP		
	XR Vector; -5' adaptor sequence: 5' GAATTCGGCAG 3' -3'		
	adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"		
BASE COUNT	117 a 97 c 88 g 165 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	159	Length:	467
Score:	56.00	Matches:	7
Percent Similarity:	66.67%	Conservative:	3
Best Local Similarity:	46.67%	Mismatches:	5
Query Match:	57.14%	Indels:	0
DB:	9	Gaps:	0
US-09-854-133-587 (1-16) x AA967175 (1-467)			
QY	2	GlnAlaAsnCysGlyIleAspPheIlePheTrpIlePheTrp 16	

Db 45 CAGCTGACATGTGCAAAATTTTTTTTGTGTTTTTTTGGGTGTTTTGG 89

RESULT 14  
BE200533  
LOCUS  
DEFINITION  
u970c08.y1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
IMAGE:1547726 5', similar to gb:X65627 M.musculus mRNA TNZ2 for p68  
RNA helicase (MOUSE);, mRNA sequence.

ACCESSION  
BE200533  
VERSION  
BE200533.1 GI:8712702  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 (bases 1 to 470)  
AUTHORS  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
MGI:953074

Seq primer: -40RP from Gibco  
High quality sequence stop: 457.  
Location/Qualifiers  
1..470  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1547726"  
/clone\_lib="Soares\_mammary\_gland\_NMLMG"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from mammary  
gland tissue from a lactating female, and was then primed  
with a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT7T3 vector. Library is normalized. Library  
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 114 a 88 c 98 g 170 t

ORIGIN

Alignment Scores:  
Pred. No.: 160 Length: 470  
Score: 56.00 Matches: 7  
Percent Similarity: 66.67% Conservative: 3  
Best Local Similarity: 46.67% Mismatches: 5  
Query Match: 57.14% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x BE200533 (1-470)

QY 2 GlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
||| ||| |||:|||||:|||||

Db 167 CAGCTGACATGTGCAAAATTTTTTTTGTGTTTTTTGGGTGTTTTGG 211

RESULT 15  
AI527491  
LOCUS  
DEFINITION  
AI527491 uf12c07.y1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
IMAGE:1511148 5', mRNA sequence.

ACCESSION  
AI527491  
VERSION  
AI527491.1 GI:4441626  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 (bases 1 to 516)  
AUTHORS  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
MGI:938000

Seq primer: -40RP from Gibco  
High quality sequence stop: 493.  
Location/Qualifiers  
1..516  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1511148"  
/clone\_lib="Soares\_mammary\_gland\_NMLMG"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from mammary  
gland tissue from a lactating female, and was then primed  
with a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT7T3 vector. Library is normalized. Library  
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 128 a 100 c 123 g 165 t

ORIGIN

Alignment Scores:  
Pred. No.: 175 Length: 516  
Score: 56.00 Matches: 7  
Percent Similarity: 66.67% Conservative: 3  
Best Local Similarity: 46.67% Mismatches: 5  
Query Match: 57.14% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x AI527491 (1-516)

QY 2 GlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
||| ||| |||:|||||:|||||

Db 421 CAGCTGACATGTGCAAAATTTTTTTTGTGTTTTTTGGGTGTTTTGG 465

Search completed: May 11, 2003, 15:28:06  
Job time : 247.549 secs



us-09-854-133-586.olig.rag

Tue May 13 12:12:42 2003

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 19:32:42 ; Search time 56.6549 Seconds  
(without alignments)  
228.141 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 97  
Sequence: 1 EVEVSRDHASLGDSFTLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0 908470

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : A\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	16	16.5	16	22	AAE13851	Human T cell epitope
2	9	9.3	188	22	ABG04555	Novel human diago
3	8	8.2	46	22	AAO04640	Human polypeptide
4	8	8.2	47	22	AAE03401	Human gene 9 encod
5	8	8.2	47	23	ABG63681	Human albumin fusi
6	8	8.2	109	22	AAU32547	Novel human secret
7	7	7.2	15	22	AAG68058	Human DNA mismatch
8	7	7.2	16	19	AAW50918	Amino acid sequenc
9	7	7.2	28	22	AAW83335	Human immune/haema
10	7	7.2	32	22	AAO05715	Human polypeptide

11	7	7.2	36	22	AAU22582	Novel human colon
12	7	7.2	36	22	AAU22555	Human digestive sy
13	7	7.2	40	22	AAO00857	Human polypeptide
14	7	7.2	41	22	AAO08040	Human polypeptide
15	7	7.2	43	22	AAO08046	Human polypeptide
16	7	7.2	47	22	AAU22244	Human cardiovascular
17	7	7.2	52	21	AAG02455	Human secreted pro
18	7	7.2	54	21	AAG00160	Human secreted pro
19	7	7.2	57	22	AAO09451	Human polypeptide
20	7	7.2	59	22	AAO10486	Human polypeptide
21	7	7.2	60	22	AAO01774	Human polypeptide
22	7	7.2	62	22	AAO02325	Human polypeptide
23	7	7.2	64	22	AAO09233	Human polypeptide
24	7	7.2	66	22	ABB03142	Human musculoskele
25	7	7.2	68	22	ABG06182	Novel human diago
26	7	7.2	80	22	AAO05335	Human polypeptide
27	7	7.2	81	23	ABB04590	Human peroxidase 9
28	7	7.2	82	22	ABG10474	Novel human diago
29	7	7.2	82	23	ABP10332	Human ORFX protein
30	7	7.2	94	23	ABB72352	Murine protein iso
31	7	7.2	96	22	AAO03592	Human polypeptide
32	7	7.2	102	22	AAO09213	Human polypeptide
33	7	7.2	102	22	AAO10463	Human polypeptide
34	7	7.2	104	21	AAG12424	Human polypeptide
35	7	7.2	109	22	ABG10473	Novel human diago
36	7	7.2	109	22	ABG26507	Human polypeptide
37	7	7.2	111	22	AAO02869	Human gene 22 enco
38	7	7.2	111	22	AAE01505	Human albumin fusi
39	7	7.2	111	23	ABG63908	Human polypeptide
40	7	7.2	114	22	AAO04586	Human polypeptide
41	7	7.2	118	22	AAO01695	Human polypeptide
42	7	7.2	120	21	AAG27487	Arabidopsis thalia
43	7	7.2	121	23	ABP43004	Human ovarian anti
44	7	7.2	124	22	ABG07519	Novel human diago
45	7	7.2	132	21	AAG54869	Arabidopsis thalia
46	7	7.2	134	20	AAU37079	Amino acid sequenc
47	7	7.2	135	22	AAG68057	Human DNA mismatch
48	7	7.2	153	21	AAU32861	Eucalyptus grandis
49	7	7.2	162	21	AAG45604	Arabidopsis thalia
50	7	7.2	168	21	AAG18723	Arabidopsis thalia
51	7	7.2	170	21	AAG18722	Arabidopsis thalia
52	7	7.2	170	22	AAU87313	Novel central nerv
53	7	7.2	170	22	AAU87603	Human novel foetal
54	7	7.2	170	22	AAU21180	Novel signal trans
55	7	7.2	170	22	AAU17251	Arabidopsis thalia
56	7	7.2	179	21	AAG18721	Novel human secret
57	7	7.2	191	22	AAU29660	Drosophila melanog
58	7	7.2	210	22	ABB59743	Human bone marrow
59	7	7.2	227	22	AAU0872	Novel human diago
60	7	7.2	253	22	ABG06957	Arabidopsis thalia
61	7	7.2	303	21	AAG47370	Human gene 24 enco
62	7	7.2	312	22	AAE01541	Novel human diago
63	7	7.2	313	22	ABG06873	Arabidopsis thalia
64	7	7.2	315	21	AAG47369	Arabidopsis thalia
65	7	7.2	318	21	AAG30396	Arabidopsis thalia

ALIGNMENTS

RESULT 1	
AAE13851	AAE13851 standard; peptide; 16 AA.
ID	AAE13851
XX	XX
AC	AAE13851;
XX	XX
DT	26-FEB-2002 (first entry)
XX	XX
DE	Human T cell epitope related to lung tumour-specific protein.
XX	XX
KW	Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW	antisense-therapy; vaccine; immune response; lung cancer;
KW	T cell epitope.

XX OS Homo sapiens.  
 XX PN WO200172295-A2.  
 XX PD 04-OCT-2001.  
 XX PF 28-MAR-2001; 2001WO-US09991.  
 XX PR 29-MAR-2000; 2000US-0538037.  
 PR 05-JUN-2000; 2000US-0588937.  
 PR 18-AUG-2000; 2000US-0640878.  
 PR 22-SEP-2000; 2000US-234517P.  
 PR 01-NOV-2000; 2000US-0704512.  
 PR 14-DEC-2000; 2000US-0738973.  
 XX XX  
 PA (CORI-) CORIXA CORP.  
 XX XX  
 PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;  
 PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
 XX WPI; 2001-639201/73.  
 XX XX  
 PT New human lung-specific polynucleotides and polypeptides for the  
 PT diagnosis and treatment of disease e.g. lung cancer -  
 XX Claim 2; Page 378; 378pp; English.  
 XX XX  
 CC The invention relates to isolated lung tumour-specific proteins and  
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
 CC their antigen-presenting cells are useful for stimulating and/or  
 CC expanding T cells specific for a tumour protein, and for inhibiting  
 CC the development of cancer. The invention also relates to a composition  
 CC useful for stimulating an immune response, and for treating cancer.  
 CC lung tumour specific oligonucleotide is useful in gene therapy and for  
 CC diagnosis, detection and treatment of lung cancer. The present sequence  
 CC is human T cell epitope related to lung tumour-specific protein.  
 XX SQ Sequence 16 AA;

Query Match 16.5%; Score 16; DB 22; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-10;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 FQANCGIDFIIIFWIFW 50  
 Db 1 FQANCGIDFIIIFWIFW 16

RESULT 2  
 ABG04555  
 ID ABG04555 standard; Protein; 188 AA.  
 XX AC ABG04555;  
 AC ABG04555;

DT 13-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #4546.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.

PN WO200175067-A2.  
 XX 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 XX PF 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX XX

PA (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS68742.  
 XX XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 XX biodiversity -  
 PS Claim 20; SEQ ID NO 34914; 103pp; English.  
 XX XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers or  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 188 AA;

Query Match 9.3%; Score 9; DB 22; Length 188;  
 Best Local Similarity 100.0%; Pred. No. 0.22;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ERKKRERK 34  
 Db 66 ERKKRERK 74

RESULT 3  
 AAO04640  
 ID AAO04640 standard; Protein; 46 AA.  
 XX AC AAO04640;  
 AC AAO04640;

DT 06-NOV-2001 (first entry)  
 DE Human polypeptide SEQ ID NO 18532.

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.  
 XX WO200164835-A2.  
 PN 07-SEP-2001.  
 PD 26-FEB-2001; 2001WO-US04927.  
 XX PF 28-FEB-2000; 2000US-0515126.  
 XX PR 18-MAY-2000; 2000US-0577409.  
 XX XX

us-09-854-133-586. olig.rag

Tue May 13 12:12:42 2003

PA (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Drmanac RT;  
PI WPI; 2001-514838/56.  
XX N-PSDB; AAI84571.  
DR Isolated nucleic acids and polypeptides, useful for preventing  
XX diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
PT Claim 20; SEQ ID NO 18532; 1399pp + Sequence Listing; English.  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
XX the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and/or  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 46 AA;  
Query Match 8.2%; Score 8; DB 22; Length 46;  
Best Local Similarity 100.0%; Pred. No. 0.67; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0;  
QY 23 RKKERKKK 30  
Db 20 RKKERKKK 27  
RESULT 4  
AAE03401  
ID AAE03401 standard; Protein; 47 AA.  
XX AAE03401;  
XX 10-AUG-2001 (first entry)  
XX Human gene 9 encoded secreted protein HKMLP68, SEQ ID NO:59.  
XX Human; secreted protein; proliferative disorder; cancer; tumour;  
XX foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder;  
KW endocrine disorder; infection; wound healing; vulnerability;  
KW cell culture; chemotaxis; food additive; gene therapy;  
KW binding partner identification.  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH 1 /label- Signal\_peptide  
FT 2..47 /label- Mature\_human\_secreted\_protein  
FT  
FT  
FT  
XX WO200132687-A1.  
PN  
XX  
XX 10-MAY-2001.  
XX  
XX  
PF 25-OCT-2000; 2000WO-US29364.  
XX 29-OCT-1999; 99US-0162238.  
PR 30-JUN-2000; 2000US-0215134.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Ruben SM, Komatsoulis GA, Duan RD, Fiscella M, Ebner R;  
PI WPI; 2001-308740/32.  
XX N-PSDB; AAD07793.  
DR Isolated nucleic acid molecule encoding a human secreted protein is  
XX used in preventing, treating or ameliorating a medical condition -  
PT Claim 11; Page 376; 398pp; English.  
XX AAD07770-AAD07794 represent cDNAs corresponding to 10 human secreted  
XX protein genes, and AAE03378-AAE03402 represent the proteins they encode.  
CC AAE03403-AAE03415 represent human secreted protein fragments or variants.  
CC The secreted proteins and their genes are useful for preventing,  
CC treating or ameliorating medical conditions, e.g., by protein or gene  
CC therapy. Pathological conditions can be diagnosed by determining the  
CC amount of the new protein in a sample or by determining the presence of  
CC mutations in the new genes. Specific uses are described for each of the  
CC 10 genes, based on the tissues in which they are most highly expressed,  
CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC abnormalities, haematopoietic disorders, diseases of the immune system,  
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
CC allergies, neurological disorders (e.g., Alzheimer's disease,  
CC Parkinson's disease), cognitive disorders, diabetes, atherosclerosis,  
CC skin disorders (e.g., psoriasis), sepsis, kidney disorders,  
CC cardiovascular disorders, angiogenic disorders, endocrine  
CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin ageing due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used  
CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein of the invention.  
XX  
SQ Sequence 47 AA;  
Query Match 8.2%; Score 8; DB 22; Length 47;  
Best Local Similarity 100.0%; Pred. No. 0.68;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 23 RKKERKKK 30  
Db 36 RKKERKKK 43  
RESULT 5  
ABG63681  
ID ABG63681 standard; Protein; 47 AA.  
XX  
XX AC ABG63681;  
XX  
XX 27-AUG-2002 (first entry)  
DT  
XX  
XX Human albumin fusion protein #356.  
XX Albumin fusion protein; therapeutic protein X; human albumin; HA;  
KW human serum albumin; HSA; cancer; reproductive disorder;  
KW digestive disorder; immune disorder; endocrine disorder;  
KW haematopoietic disorder; neural disorder; connective disorder;  
KW cytostatic; antiinfertility; antiinflammatory; antiulcer;

immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;  
 neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
 osteopathic; antiarthritic.  
 Homo sapiens.  
 Synthetic.  
 WO200177137-A1.  
 18-OCT-2001.  
 12-APR-2001; 2001WO-US11988.  
 12-APR-2000; 2000US-229358P.  
 25-APR-2000; 2000US-199384P.  
 21-DEC-2000; 2000US-256931P.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Haseltine WA;  
 WPI; 2002-010886/01.  
 New fusion protein for treating disease e.g. diabetes comprises an  
 albumin fused to a therapeutic protein.  
 Claim 1; Page 789; 2102pp; English.  
 The present invention relates to albumin fusion proteins comprising a  
 therapeutic protein X and human albumin (HA, also known as human serum  
 albumin, HSA). The proteins are useful for treating a disease or  
 disorder that may be modulated by therapeutic protein X. The albumin  
 extends the shelf-life of protein X, and may increase its biological  
 activity. The protein is useful for treating and  
 diagnosing disorders such as cancer, reproductive disorders, digestive  
 disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders  
 (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders  
 (e.g. diabetes), haematopoietic disorders, neural disorders  
 (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,  
 encephalomyelitis, meningitis, schizophrenia), and connective disorders  
 (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin  
 fusion proteins of the invention.  
 Sequence 47 AA;  
 Query Match 8.2%; Score 8; DB 23; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 0.68;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKKKKK 30  
 Db 36 RKKKKKK 43  
 RESULT 6  
 AAU32547  
 ID AAU32547 standard; Protein; 109 AA.  
 AC AAU32547;  
 DT 18-DEC-2001 (first entry)  
 DE Novel human secreted protein #3038.  
 KW Human; vaccination; gene therapy; nutritional supplement;  
 stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
 immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
 Homo sapiens.  
 WO200179449-A2.  
 25-OCT-2001.

16-APR-2001; 2001WO-US08656.  
 18-APR-2000; 2000US-0552929.  
 26-JAN-2001; 2001US-0770160.  
 (HYSE-) HYSEQ INC.  
 Tang YT, Liu C, Drmanac RT;  
 WPI; 2001-611725/70.  
 Nucleic acids encoding a range of human polypeptides, useful in genetic  
 vaccination, testing and therapy.  
 Claim 20; Page 629; 765pp; English.  
 The invention relates to novel human secreted polypeptides. The  
 polypeptides and antibodies to the polypeptides are useful for  
 determining the presence of or predisposition to a disease associated  
 with altered levels of polypeptide. The polypeptides are also useful for  
 identifying agents (agonists and antagonists) that bind to them. Cells  
 expressing the proteins are useful for identifying a therapeutic agent  
 for use in treatment of a pathology related to aberrant expression or  
 physiological interactions of the polypeptide. Vectors comprising  
 the nucleic acids encoding the polypeptides and cells genetically  
 engineered to express them are also useful for producing the proteins.  
 The proteins are useful in genetic vaccination, testing and  
 therapy, and can be used as nutritional supplements. They may be used to  
 increase stem cell proliferation; to regulate haematopoiesis; and in  
 bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
 immune suppression and/or stimulation; as anti-inflammatory agents; and  
 in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid  
 sequences of novel human secreted proteins of the invention.  
 Sequence 109 AA;  
 Query Match 8.2%; Score 8; DB 22; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 1.5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 28 KKKRERKF 35  
 Db 83 KKKRERKF 90  
 RESULT 7  
 AAG68058  
 ID AAG68058 standard; Peptide; 15 AA.  
 AC AAG68058;  
 DT 13-DEC-2001 (first entry)  
 DE Human DNA mismatch repair protein 15 N-terminal peptide SEQ ID NO:7.  
 KW Human; DNA mismatch repair protein 15; cytostatic; virucidal;  
 immunomodulatory; antiinflammatory; haemostatic; anti-HIV; gene therapy;  
 diagnosis; malignant tumour; haemopathy; human immunodeficiency virus;  
 HIV infection; immunological disease; inflammation.  
 Homo sapiens.  
 WO200170958-A1.  
 27-SEP-2001.  
 19-MAR-2001; 2001WO-CN00356.  
 22-MAR-2000; 2000CN-0115043.  
 (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.



us-09-854-133-586.olg.rag

Tue May 13 12:12:42 2003

PI Mao Y, Xie Y;  
XX WPI; 2001-602782/68.  
DR New human DNA mismatch repair protein 15 for diagnosing and treating  
XX malignant tumor, hemopathy, human immunodeficiency virus infection,  
PT immunological diseases and inflammations  
PT Example 5; Page 14; 34pp; Chinese.  
XX The present invention describes human DNA mismatch repair protein 15 (I).  
PS (I) has cytostatic, virucidal, immunomodulatory, antiinflammatory,  
XX haemostatic and anti-HIV activities. The polynucleotide (II) encoding  
CC (I) can be used in gene therapy. (I) and (II) can be used in the  
CC diagnosis and treatment of malignant tumour, haemopathy, human  
CC immunodeficiency virus (HIV) infection, immunological diseases and  
CC various inflammations. The present sequence represents the human DNA  
CC mismatch repair protein 15 N-terminal peptide, which is used in an  
CC example from the present invention.  
XX  
SQ Sequence 15 AA;

Query Match 7.2%; Score 7; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 RKKERKK 29  
Db 9 RKKERKK 15

RESULT 8  
AAW50918  
ID AAW50918 standard; peptide; 16 AA.

XX AAW50918;  
XX 09-SEP-1998 (first entry)  
XX Amino acid sequence of a human epithelial peptide.  
DE Human epithelial peptide; marker; cancer; probe; hybridisation;  
XX primer; amplification; lung; liver; kidney; breast; prostate;  
KW melanoma; myeloma; antibody.  
XX Homo sapiens.

XX WO9814469-A2.  
XX 09-APR-1998.  
XX 02-OCT-1997; 97WO-US17714.  
XX 02-OCT-1996; 96US-0725027.

XX (UYJO ) UNIV JOHNS HOPKINS.  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX Mulshine JL, Tockman MS;  
XX WPI; 1998-240016/21.

XX New isolated epithelial protein as early marker of cancer - useful  
PT in computer-assisted methods of diagnosis based on discriminant  
PT analysis of optical images of cells  
XX Claim 2; Page 10; 159; English.

XX This is the amino acid sequence of the human epithelial peptide, used  
CC in the method of the invention as early markers for cancer. Probes  
CC and primers that hybridise to or amplify these peptides are used to  
CC diagnose precancerous states, e.g. of lung, liver, kidney, breast,  
CC prostate, head or neck, melanoma or myeloma, or to determine

CC susceptibility to these conditions and for monitoring treatment.  
CC precancer is also indicated by detecting post-translational  
CC modification of the epithelial peptide which is a marker of epithelial  
CC cell transformation. Antibodies are potentially useful for diagnosis  
CC and treatment of cancer.  
XX Sequence 16 AA;  
SQ Query Match 7.2%; Score 7; DB 19; Length 16;  
Best Local Similarity 100.0%; Pred. No. 2.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 ERKKKRE 32  
Db 10 ERKKKRE 16

RESULT 9  
AAM83335  
ID AAM83335 standard; Protein; 28 AA.

XX AAM83335;  
XX 07-NOV-2001 (first entry)  
XX Human immune/haematopoietic antigen SEQ ID NO:10928.  
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis.  
XX Homo sapiens.

XX WO200157182-A2.  
XX 09-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0231958.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0234998.  
PR 27-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 29-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2000US-0254097.  
XX 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.  
N-PSDB; AAK56116.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
useful for preventing, diagnosing and/or treating cancers and  
metastasis

Claim 11; SEQ ID NO 10928; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
activity, and can be used in gene therapy and vaccine production. (I)  
proteins and polynucleotides may be used in the prevention, diagnosis and  
treatment of diseases associated with inappropriate (I) expression. For  
example, they may be used to treat disorders associated with decreased  
expression by rectifying mutations or deletions in a patient's genome  
that affect the activity of (I) by expressing inactive proteins or to  
supplement the patients own production of (I). Additionally, (I)  
polynucleotides may be used to produce the secreted (I), by inserting  
the nucleic acids into a host cell and culturing the cell to express the  
protein. (I) proteins and polynucleotides may be used to prevent,  
diagnose and treat immune/haematopoietic-related diseases, especially  
cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
to AAK87694 represent human immune/haematopoietic antigen genomic  
sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
represent sequences used in the exemplification of the present invention.

Sequence 28 AA;

Query Match

Best Local Similarity 7.2%; Score 7; DB 22; Length 28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKKKKK 30  
|||||

Db 12 KKKKKK 18

RESULT 10  
AAO05715



PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234223.  
PR 25-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 26-SEP-2000; 2000US-0234998.  
PR 27-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.

PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI; 2001-465567/50.  
DR N-PSDB; AAS39462.  
XX  
PT Isolated polypeptide for treating, preventing and/or prognosing  
PT disorders related to the colon including colon cancers and also for  
PT testing and detection e.g. diagnosis -  
XX  
PS Claim 11; SEQ ID No 359; 562pp; English.  
XX  
CC The present invention relates to the isolation of novel human colon  
CC associated polypeptides, and the cDNA (AAS39348-AAS39581) and genomic  
CC sequences encoding for them. The sequences of the invention are useful  
CC in the diagnosis, treatment, prevention and/or prognosis of disorders  
CC of the colon including colon cancer, congenital abnormalities  
CC (e.g. atresia and stenosis), bacterial and viral infections,  
CC inflammatory bowel disease (IBD), neoplastic cell disorders,  
CC (e.g. polyps and adenomas, intestinal inflammatory disorders, colitis,  
CC colonic inflammation, diarrhoea and dysentery, malabsorption syndromes  
CC (e.g. lactose intolerance), intestinal obstruction and sigmoid diseases.  
CC The polynucleotide sequences of the invention can also be used in gene  
CC therapy. AAU22468-AAU22701 represent the novel human colon associated  
CC polypeptides of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 36 AA;  
  
Query Match 7.2%; Score 7; DB 22; Length 36;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 24 KKERKKK 30  
Db 7 KKERKKK 13  
  
RESULT 12  
AAM92555  
ID AAM92555 standard; Protein; 36 AA.  
XX  
AC AAM92555;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human digestive system antigen SEQ ID NO: 1904.  
XX  
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
XX digestive system disorder; Meckel's diverticulum.  
OS Homo sapiens.  
XX  
PN WO200155314-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01324.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.



us-09-854-133-586.olg.rag

Tue May 13 12:12:42 2003

PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0226279.  
PR 18-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 22-AUG-2000; 2000US-0227009.  
PR 23-AUG-2000; 2000US-0228924.  
PR 30-AUG-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 01-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 05-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 06-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 14-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 21-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
PI WPI; 2001-502630/55.  
DR N-PSDB; AAK88328.  
DR  
XX  
PT Polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases -  
XX  
PS Claim 11; SEQ ID NO 1904; 986pp; English.  
XX  
CC The present invention provides the protein and coding sequences of a

CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a digestive system antigen of  
CC the invention.  
XX  
SQ Sequence 36 AA;

Query Match 7.2%; Score 7; DB 22; Length 36;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKKKKK 30  
Db 7 KKKKKK 13

RESULT 13  
AAO00857  
ID AAO00857 standard; Protein; 40 AA.  
XX  
AC AAO00857;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 14749.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

XX WO200164835-A2.

PN 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR N-PSDB; AAI80788.

XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -

PS Claim 20; SEQ ID NO 14749; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 40 AA;

Query Match 7.2%; Score 7; DB 22; Length 40;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQT 20  
Db 22 SETLSQT 28

RESULT 14  
AAO08040  
ID AAO08040 standard; Protein; 41 AA.  
XX  
AC AAO08040;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 21932.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

XX WO200164835-A2.

PN 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR N-PSDB; AAI87971.

XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -

PS Claim 20; SEQ ID NO 21932; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 41 AA;

Query Match 7.2%; Score 7; DB 22; Length 41;  
Best Local Similarity 100.0%; Pred. No. 6.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKKKK 29  
Db 16 RKKKKK 22

us-09-854-133-586.olig.rag

Tue May 13 12:12:42 2003

```

RESULT 15
AAO08046
ID AAO08046 standard; Protein; 43 AA.
XX
AC AAO08046;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 21938.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
PI WPI; 2001-514838/56.
XX
DR N-PSDB; AAI87977.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders.
XX
PS Claim 20; SEQ ID NO 21938; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 43 AA;

Query Match 7.2%; Score 7; DB 22; Length 43;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKKKK 29
Db 13 RKKKKK 19

RESULT 16
AAU22244
ID AAU22244 standard; Protein; 47 AA.
XX
AC AAU22244;
XX
```

```

DT 17-DEC-2001 (first entry)
XX
DE Human cardiovascular system antigen polypeptide SEQ ID No 1018.
XX
KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; tissue regeneration;
XX anti-infertility.
XX
OS Homo sapiens.
XX
PN WO200155321-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01340.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
```

PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI; 2001-451930/48.  
DR N-PSDB; AAS35518.  
XX  
PT  
PT  
PT  
XX  
PS  
XX

New cardiovascular system related polynucleotides and polypeptides,  
useful for diagnosing, treating and/or preventing disorders of the  
cardiovascular system -

Claim 11; SEQ ID No 1018; 674pp; English.

Sequences AAU21852-AAU22466 represent the cardiovascular system antigen  
polypeptides of the invention. Cardiovascular system antigens and their  
associated polynucleotides are useful in the diagnosis, treatment and  
prevention of various types of disorders in e.g. humans, mice, rabbits,  
goats, horses, cats, dogs, chickens or sheep. A pathological condition  
can be determined by detecting the presence or absence of a mutation in a  
cardiovascular system antigen polynucleotide. The treatable disorders  
include autoimmune diseases such as rheumatoid arthritis,  
hyperproliferative disorders such as neoplasms of the breast or liver,  
cardiovascular disorders such as cardiac arrest, cerebrovascular  
disorders such as cerebral ischaemia, nervous system disorders such as  
Alzheimer's disease, infections caused by bacteria, viruses and fungi,  
ocular disorders such as corneal infection, endocrine disorders such as  
premature labour and infertility, gastrointestinal disorders such as  
Crohn's disease, renal disorders such as glomerulonephritis and  
respiratory disorders such as asthma and pleurisy. The polypeptides can  
also be used to aid wound healing, to prevent skin aging due to sunburn,  
to maintain organs before transplantation, to regenerate tissues and in  
chemotaxis.  
Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 7.2%; Score 7; DB 22; Length 47;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 25 KERKKR 31  
Db 14 KERKKR 20  
|||||

RESULT 17  
AAG02455  
ID AAG02455 standard; Protein; 52 AA.  
XX  
AC AAG02455;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein, SEQ ID NO: 6536.  
XX  
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping.



Tue May 13 12:12:42 2003

OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI; 2000-500381/45.  
 DR N-PSDB; AAC02461.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 PT  
 XX  
 PS Claim 13; SEQ ID 6536; 7lpp + CD-ROM; English.  
 XX  
 CC The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.  
 XX  
 SQ Sequence 52 AA;  
 Query Match 7.2%; Score 7; DB 21; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 8.6;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKERKK 29  
 Db 18 RKKERKK 24  
 RESULT 18  
 AAG00160  
 ID AAG00160 standard; Protein; 54 AA.  
 XX  
 AC AAG00160;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein, SEQ ID NO: 4241.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GEST ) GENSET.  
 XX

PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI; 2000-500381/45.  
 DR N-PSDB; AAC00166.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 PT  
 XX  
 PS Claim 13; SEQ ID 4241; 7lpp + CD-ROM; English.  
 XX  
 CC The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.  
 XX  
 SQ Sequence 54 AA;  
 Query Match 7.2%; Score 7; DB 21; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 8.9;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 SLGDSET 16  
 Db 45 SLGDSET 51  
 RESULT 19  
 AAO09451  
 ID AAO09451 standard; Protein; 57 AA.  
 XX  
 AC AAO09451;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 23343.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200164835-A2.  
 XX  
 PD 07-SEP-2001.  
 XX  
 DE 26-FEB-2001; 2001WO-US04927.  
 XX  
 PF 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 PR  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Liu C, Drmanac RT;  
 PI  
 XX WPI; 2001-514838/56.  
 DR N-PSDB; AAI89382.  
 XX  
 PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 XX

PS Claim 20; SEQ ID NO 23343; 1399pp + Sequence Listing; English.  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and/or  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 57 AA;

Query Match 7.2%; Score 7; DB 22; Length 57;  
Best Local Similarity 100.0%; Pred. No. 9.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERK 29  
Db 4 RKKERK 10

RESULT 20  
AAO10486  
ID AAO10486 standard; Protein; 59 AA.

XX AAO10486;  
AC  
XX  
DT 06-NOV-2001 (first entry)  
XX Human polypeptide SEQ ID NO 24378.  
DE  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.

XX WO200164835-A2.  
PN  
XX  
PD 07-SEP-2001.  
XX

XX 26-FEB-2001; 2001WO-US04927.  
PF  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

DR WPI; 2001-514838/56.  
DR N-PSDB; AAI90417.

XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders.

PS Claim 20; SEQ ID NO 24378; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 59 AA;

Query Match 7.2%; Score 7; DB 22; Length 59;  
Best Local Similarity 100.0%; Pred. No. 9.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERK 29  
Db 46 RKKERK 52

RESULT 21  
AAO01774  
ID AAO01774 standard; Protein; 60 AA.

XX AAO01774;  
AC  
XX

DT 06-NOV-2001 (first entry)  
XX

DE Human polypeptide SEQ ID NO 15666.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

DR WPI; 2001-514838/56.  
DR N-PSDB; AAI81705.

XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders.

PS Claim 20; SEQ ID NO 15666; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

Tue May 13 12:12:42 2003

```
XX SQ Sequence 60 AA;
Query Match 7.2%; Score 7; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29
Db 35 RKKERKK 41
|||||

RESULT 22
AAO02325
ID AAO02325 standard; Protein; 62 AA.
XX
AC AAO02325;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 16217.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
DE Human polypeptide SEQ ID NO 16217.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
XX
DR N-PSDB; AAI82256.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 20; SEQ ID NO 16217; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 62 AA;
Query Match 7.2%; Score 7; DB 22; Length 62;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29
Db 35 RKKERKK 41
|||||

RESULT 24
ABB03142
ID ABB03142 standard; Protein; 66 AA.
XX
```

```
Db 50 RKKERKK 56
|||||

RESULT 23
AAO09233
ID AAO09233 standard; Protein; 64 AA.
XX
AC AAO09233;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 23125.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
DE Human polypeptide SEQ ID NO 23125.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
XX
DR N-PSDB; AAI89164.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 20; SEQ ID NO 23125; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 64 AA;
Query Match 7.2%; Score 7; DB 22; Length 64;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29
Db 16 RKKERKK 22
|||||

RESULT 24
ABB03142
ID ABB03142 standard; Protein; 66 AA.
XX
```

AC ABB03142;  
XX  
DT 08-JAN-2002 (first entry)  
XX  
DE Human musculoskeletal system related polypeptide SEQ ID NO 1089.  
XX  
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system.  
XX  
OS Homo sapiens.  
XX  
PN WO200155367-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01338.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 20-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.



us-09-854-133-586.olg.rag

Tue May 13 12:12:42 2003

```

PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-451937/48.
DR N-PSDB; AAL34724.
XX
XX Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the musculoskeletal system including
PT musculoskeletal cancers and also for testing and detection e.g.
PT diagnosis -
XX
XX Claim 11; SEQ ID NO 1089; 781pp + Sequence Listing; English.
PS
XX
XX The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
CC for preventing, treating or ameliorating medical conditions e.g. by
CC protein or gene therapy. The genes are isolated from a range of human
CC tissues disclosed in the specification. The nucleic acids, proteins,
CC antibodies and (ant)agonists are useful in the diagnosis, treatment
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischaemia; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 66 AA;
SQ
Query Match 7.2%; Score 7; DB 22; Length 66;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 24 KKERKKK 30
Db 35 KKERKKK 41
RESULT 25
ABG06182
ID ABG06182 standard; Protein; 68 AA.
XX
XX ABG06182;
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #6173.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN

XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR N-PSDB; AAS70369.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID NO 36541; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant products of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 68 AA;
SQ
Query Match 7.2%; Score 7; DB 22; Length 68;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 24 KKERKKK 30
Db 45 KKERKKK 51
RESULT 26
AAO05335
ID AAO05335 standard; Protein; 80 AA.
XX
XX AAO05335;
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human polypeptide SEQ ID NO 19227.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
XX Homo sapiens.
OS
XX WO200164835-A2.
PN

```

XX PD 07-SEP-2001.  
 XX PF 26-FEB-2001; 2001WO-US04927.  
 XX PR 28-FEB-2000; 2000US-0515126.  
 XX PR 18-MAY-2000; 2000US-0577409.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Drmanac RT;  
 XX DR WPI; 2001-514838/56.  
 XX DR N-PSDB; AAI85266.  
 XX PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders.  
 XX SQ Claim 20; SEQ ID NO 19227; 1399pp + Sequence Listing; English.  
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 80 AA;  
 Query Match  
 Best Local Similarity 7.2%; Score 7; DB 22; Length 80;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKERKK 29  
 Db 52 RKKERKK 58  
 RESULT 27  
 ABB04590  
 ID ABB04590 standard; Protein; 81 AA.  
 XX AC ABB04590;  
 XX DT 21-MAR-2002 (first entry)  
 XX DE Human peroxidase 9.  
 KW Human; peroxidase 9; cancer; haemopathy; HIV infection; phlogosis;  
 KW cytostatic; haemostatic; virucide; immunomodulatory; antiinflammatory;  
 KW immune disease; gene therapy; protein; enzyme.  
 XX OS Homo sapiens.  
 XX WO200192317-A1.  
 PN WO200192317-A1.  
 PD 06-DEC-2001.  
 XX PF 21-MAY-2001; 2001WO-CN00844.  
 XX PR 24-MAY-2000; 2000CN-0115848.  
 XX PA (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.

PI Mao Y, Xie Y;  
 XX WPI; 2002-090029/12.  
 DR N-PSDB; ABA05891.  
 XX HOMO peroxidase 9 and encoding polynucleotide, used in diagnosis and  
 PT treatment of malignant tumors, hemopathy, human immunodeficiency virus  
 PT infection, immunological diseases and inflammation -  
 XX Claim 1; Page 29; 37pp; Chinese.  
 PS The present invention provides the protein and coding sequences of human  
 CC peroxidase 9. The sequences can be used in the treatment of cancer,  
 CC haemopathy, HIV infection, phlogosis and immune diseases. The present  
 CC sequence is the protein of the invention.  
 XX SQ Sequence 81 AA;  
 Query Match  
 Best Local Similarity 7.2%; Score 7; DB 23; Length 81;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKERKK 29  
 Db 21 RKKERKK 27  
 RESULT 28  
 ABG10474  
 ID ABG10474 standard; Protein; 82 AA.  
 XX AC ABG10474;  
 XX DT 13-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #10465.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX WO200175067-A2.  
 PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US08631.  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS74661.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX Claim 20; SEQ ID No 40833; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or

Tue May 13 12:12:42 2003

quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 82 AA;  
Query Match 7.2%; Score 7; DB 22; Length 82;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKKKK 29  
Db 17 RKKKKK 23

RESULT 29  
ABP10332  
ID ABP10332 standard; Protein; 82 AA.

XX  
AC ABP10332;  
XX  
DT 25-JUN-2002 (first entry)  
XX  
DE Human ORFX protein sequence SEQ ID NO:20646.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.

XX Homo sapiens.  
XX WO200192523-A2.  
XX  
XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US10836.  
XX  
XX 30-MAY-2000; 2000US-206132P.  
XX  
XX 29-AUG-2000; 2000US-228716P.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach MD;

XX WPI; 2002-106308/14.  
XX N-PSDB; ABN26084.

XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders -

XX Disclosure; SEQ ID 20646; 1037pp; English.

XX The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated

disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, or periodontal disease, and for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues and conditions resulting from systemic cytokine damage.

QY 24 KKKKKK 30  
Db 18 KKKKKK 24

Query Match 7.2%; Score 7; DB 23; Length 82;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKKKKK 30  
Db 18 KKKKKK 24

RESULT 30  
ABB72352  
ID ABB72352 standard; Protein; 94 AA.

XX  
AC ABB72352;  
XX  
DT 04-APR-2002 (first entry)

XX Murine protein isolated from skin cells SEQ ID NO: 676.

XX Human; rat; mouse; skin cell; skin wound; cancer; growth defect;  
KW developmental defect; inflammatory disease; dermatological; vulnary;  
KW immunomodulator; anti-inflammatory; cytostatic; neuroprotective.

XX Mus sp.

XX WO200190357-A1.

XX 29-NOV-2001.

XX 24-MAY-2001; 2001WO-NZ00099.

XX 24-MAY-2000; 2000US-206650P.

XX 25-JUL-2000; 2000US-221232P.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG, Kumble KD;

XX WPI; 2002-122020/16.  
XX N-PSDB; ABL35041.

XX New polynucleotides and polypeptides encoded by the polynucleotides  
PT isolated from skin cells, useful for treating skin wounds, cancers,  
PT growth and developmental defects, inflammatory diseases, or for  
PT modulating immune responses -

XX Claim 4; Page 432; 466pp; English.

XX The present invention provides the protein and coding sequences of cDNAs  
CC isolated from human, murine and rat skin cell libraries. The sequences

CC can be used in the development of therapeutic agents useful in the  
CC treatment of skin diseases, including skin wounds, cancer, growth  
CC defects, developmental defects and inflammatory diseases. The proteins  
CC have important roles in the induction of hair growth, cell proliferation  
CC and cell-cell interaction, in maintaining tissue integrity, in wound  
CC healing and in modulating immune responses. The present sequence is a  
CC polypeptide of the invention.

XX SQ Sequence 94 AA;

Query Match 7.2%; Score 7; DB 23; Length 94;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERK 29  
Db 35 RKKERK 41  
|||||

RESULT 31  
AAO03592  
ID AAO03592 standard; Protein; 96 AA.

XX AC AAO03592;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 17484.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-514838/56.

XX DR N-PSDB; AAI83523.

XX PT Isolated nucleic acids and polypeptides, useful for preventing  
XX diagnosing and treating e.g. leukaemia, inflammation and immune  
XX disorders -

XX PS Claim 20; SEQ ID NO 17484; 1399pp + Sequence Listing; English.

XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
XX production of other cytokines in other cell populations. The  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX activin/inhibin activity and may be useful in the diagnosis and/or  
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX inflammation.

XX CC Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 96 AA;

Query Match 7.2%; Score 7; DB 22; Length 96;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERK 29  
Db 24 RKKERK 30  
|||||

RESULT 32  
AAO09213  
ID AAO09213 standard; Protein; 102 AA.

XX AC AAO09213;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 23105.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-514838/56.

XX DR N-PSDB; AAI89144.

XX PT Isolated nucleic acids and polypeptides, useful for preventing  
XX diagnosing and treating e.g. leukaemia, inflammation and immune  
XX disorders -

XX PS Claim 20; SEQ ID NO 23105; 1399pp + Sequence Listing; English.

XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
XX production of other cytokines in other cell populations. The  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX activin/inhibin activity and may be useful in the diagnosis and/or  
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX inflammation.

XX CC Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 102 AA;

Query Match 7.2%; Score 7; DB 22; Length 102;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERK 29  
|||||



Tue May 13 12:12:42 2003

Db 29 RKKRKK 35

RESULT 33  
AAO10463  
ID AAO10463 standard; Protein; 102 AA.  
XX  
AC AAO10463;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 24355.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI; 2001-514838/56.  
DR N-PSDB; AAI90394.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders.  
XX  
PS Claim 20; SEQ ID NO 24355; 1399pp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AAI9941-AAI93841) and  
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 102 AA;  
  
Query Match 7.2%; Score 7; DB 22; Length 102;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 24 KKKKKK 30  
DB 57 KKKKKK 63  
  
RESULT 34  
AAG12424  
ID AAG12424 standard; Protein; 104 AA.  
XX  
AC AAG12424;  
  
XX 17-OCT-2000 (first entry)  
DT Zea mays protein fragment SEQ ID NO: 11530.  
DE Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.  
XX  
OS Zea mays subsp. mays.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148584.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 25-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0157865.  
PR 08-OCT-1999; 99US-0158029.  
PR 12-OCT-1999; 99US-0158232.  
PR 13-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 7.28; Score 7; DB 21; Length 104;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 KKKKKK 30  
Db 37 KKKKKK 43

RESULT 35  
ABG10473  
ID ABG10473 standard; Protein; 109 AA.  
XX  
AC ABG10473;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #10464.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

us-09-854-133-586.olog.rag

Tue May 13 12:12:42 2003

XX	WO200175067-A2.	XX	WO200175067-A2.
OS	Homo sapiens.	PN	
XX		XX	
XX		PD	
PN	WO200175067-A2.	XX	11-OCT-2001.
XX		XX	30-MAR-2001; 2001WO-US08631.
PD	11-OCT-2001.	PF	
XX		XX	31-MAR-2000; 2000US-0540217.
XX		PR	23-AUG-2000; 2000US-0649167.
PF		XX	(HYSE-) HYSEQ INC.
XX		PA	
XX		XX	Drmanac RT, Liu C, Tang YT;
PA	(HYSE-) HYSEQ INC.	PI	
XX		XX	WPI; 2001-639362/73.
XX		DR	N-PSDB; AAS90694.
PI	Drmanac RT, Liu C, Tang YT;	XX	
XX		XX	New isolated polynucleotide and encoded polypeptides, useful in
DR	WPI; 2001-639362/73.	PT	diagnostics, forensics, gene mapping, identification of mutations
DR	N-PSDB; AAS74660.	PT	responsible for genetic disorders or other traits and to assess
XX		PT	biodiversity
XX		PS	Claim 20; SEQ ID No 40832; 103pp; English.
PS	Claim 20; SEQ ID No 40832; 103pp; English.	XX	
XX		CC	The invention relates to isolated polynucleotide (I) and
CC		CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC		CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC		CC	and gene mapping, and in recombinant production of (II). The
CC		CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC		CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC		CC	to restore normal activity of (II) or to treat disease states involving
CC		CC	(II). (II) is useful for generating antibodies against it, detecting or
CC		CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC		CC	a food supplement. (II) and its binding partners are useful in medical
CC		CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC		CC	disorders involving aberrant protein expression or biological activity.
CC		CC	The polypeptide and polynucleotide sequences have applications in
CC		CC	diagnostics, forensics, gene mapping, identification of mutations
CC		CC	responsible for genetic disorders or other traits to assess biodiversity
CC		CC	and to produce other types of data and products dependent on DNA and
CC		CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC		CC	diagnostic amino acid sequences of the invention.
CC		CC	Note: The sequence data for this patent did not appear in the printed
CC		CC	specification, but was obtained in electronic format directly from WIPO
CC		CC	at ftp.wipo.int/pub/published_pct_sequences.
XX		XX	Sequence 109 AA;
SQ	Sequence 109 AA;	SQ	
			Query Match 7.2%; Score 7; DB 22; Length 109;
			Best Local Similarity 100.0%; Pred. No. 17;
			Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	23 RKKERKK 29	QY	23 RKKERKK 29
DB	38 RKKERKK 44	DB	82 RKKERKK 88
			RESULT 37
			AAO02869
			ID AAO02869 standard; Protein; 109 AA.
			XX
			AC AAO02869;
			XX
			XX
			DT 06-NOV-2001 (first entry)
			XX
			XX
			DE Human polypeptide SEQ ID NO 16761.
			XX
			XX
			KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
			KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
			KW tissue growth factor; immunomodulatory; cancer; leukaemia;
			KW nervous system disorders; arthritis; inflammation.
			XX
			OS Homo sapiens.

XX WO200164835-A2.  
PN 07-SEP-2001.  
XX 26-FEB-2001; 2001WO-US04927.  
PD 28-FEB-2000; 2000US-0515126.  
XX 18-MAY-2000; 2000US-0577409.  
XX (HYSE-) HYSEQ INC.  
PA Tang YT, Liu C, Drmanac RT;  
XX WPI; 2001-514838/56.  
PI N-PSDB; AAI82800.  
DR Isolated nucleic acids and polypeptides, useful for preventing  
DR diagnosing and treating e.g. leukaemia, inflammation and immune  
XX disorders -  
PT Claim 20; SEQ ID NO 16761; 1399pp + Sequence Listing; English.  
PT The invention relates to human polynucleotides (AAI79941-AAI93841) and  
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin growth factor activity, immunomodulatory activity and  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 111 AA;  
Query Match 7.2%; Score 7; DB 22; Length 111;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 23 RKKERKK 29  
Db 19 RKKERKK 25  
RESULT 38  
AAE01505  
ID AAE01505 standard; Protein; 111 AA.  
XX AC AAE01505;  
XX 17-JUL-2001 (first entry)  
DT Human gene 22 encoded secreted protein HE9SE18, SEQ ID NO:162.  
DE Human; secreted protein; proliferative disorder; cancer; tumour;  
XX foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; kidney disease;  
KW cardiovascular disorder; angiotensin-related disorder; kidney disorder;  
KW gastrointestinal disorder; infection; pregnancy-related disorder;  
KW endocrine disorder; wound healing; vulnerability;  
KW cell culture; chemotaxis; food additive; gene therapy;  
XX binding partner identification.  
OS Homo sapiens.  
XX

FH Key Location/Qualifiers  
FT Peptide I  
FT Protein /label- Signal\_peptide  
FT 2..111  
XX /label- Human\_mature\_secreted\_protein  
PN WO200134626-A1.  
XX 17-MAY-2001.  
PD 01-NOV-2000; 2000WO-US30045.  
XX 05-NOV-1999; 99US-0163581.  
PR 30-JUN-2000; 2000US-0215133.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Ruben SM, Komatsoulis GA, Moore PA, Birse CE, NI J;  
PI WPI; 2001-308778/32.  
XX N-PSDB; AAD05371.  
DR New nucleic acid molecules encoding 28 human secreted proteins for  
XX diagnosing, preventing, treating or ameliorating medical conditions and  
XX used as food additives or preservatives -  
PS Claim 11; Page 527-528; 562pp; English.  
XX AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted  
CC protein genes, and AAE01436-AAE01513 represent the proteins they encode.  
CC AAE01514-AAE01544 represent human secreted protein fragments or variants.  
CC The genes and their secreted proteins are useful for preventing,  
CC treating or ameliorating medical conditions, e.g., by protein or gene  
CC therapy. Pathological conditions can be diagnosed by determining the  
CC amount of the new protein in a sample or by determining the presence of  
CC mutations in the new genes. Specific uses are described for each of the  
CC 28 genes, based on the tissues in which they are most highly expressed,  
CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC abnormalities, haematopoietic disorders, diseases of the immune system,  
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
CC allergies, neurological disorders (e.g., Alzheimer's disease,  
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
CC cardiovascular disorders, angiogenic disorders, kidney disorders,  
CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used  
CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g. radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
XX secreted protein of the invention.  
SQ Sequence 111 AA;  
Query Match 7.2%; Score 7; DB 22; Length 111;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 23 RKKERKK 29  
Db 92 RKKERKK 98  
RESULT 39  
ABG63908  
ID ABG63908 standard; Protein; 111 AA.  
XX



Tue May 13 12:12:42 2003

AC ABG63908;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Human albumin fusion protein #583.  
XX  
XX Albumin fusion protein; therapeutic protein X; human albumin; HA;  
KW human serum albumin; HSA; cancer; reproductive disorder;  
KW digestive disorder; immune disorder; endocrine disorder;  
KW haematopoietic disorder; neural disorder; connective disorder;  
KW cytostatic; antiinfertility; antiinflammatory; antiulcer;  
KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;  
KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
KW osteopathic; antiarthritic.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO200177137-A1.  
XX  
PD 18-OCT-2001.  
XX  
PF 12-APR-2001; 2001WO-US11988.  
XX  
PR 12-APR-2000; 2000US-29358P.  
PR 25-APR-2000; 2000US-199384P.  
PR 21-DEC-2000; 2000US-256931P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX Rosen CA, Haseltine WA;  
PI  
XX WPI; 2002-010886/01.  
DR  
XX New fusion protein for treating disease e.g. diabetes comprises an  
XX albumin fused to a therapeutic protein -  
PT  
XX  
XX Claim 1; Page 899-900; 2102pp; English.  
PS  
XX The present invention relates to albumin fusion proteins comprising a  
CC therapeutic protein X and human albumin (HA, also known as human serum  
CC albumin, HSA). The proteins are useful for treating a disease or  
CC disorder that may be modulated by therapeutic protein X. The albumin  
CC extends the shelf-life of protein X, and may increase its biological  
CC in vitro/in vivo activity. The protein is useful for treating and  
CC diagnosing disorders such as cancer, reproductive disorders, digestive  
CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders  
CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders  
CC (e.g. diabetes), haematopoietic disorders, neural disorders  
CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,  
CC encephalomyelitis, meningitis, schizophrenia), and connective disorders  
CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin  
CC fusion proteins of the invention.  
XX  
SQ Sequence 111 AA;  
Query Match 7.2%; Score 7; DB 23; Length 111;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 23 RKKKKK 29  
Db 92 RKKKKK 98  
RESULT 40  
AAO04586  
ID AAO04586 standard; Protein; 114 AA.  
XX  
AC AAO04586;  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX

DE Human polypeptide SEQ ID NO 18478.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-514838/56.  
DR N-PSDB; AAI84517.  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
XX Claim 20; SEQ ID NO 18478; 1399pp + Sequence Listing; English.  
PS  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 114 AA;  
Query Match 7.2%; Score 7; DB 22; Length 114;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 24 KKKKKK 30  
Db 81 KKKKKK 87  
RESULT 41  
AAO01695  
ID AAO01695 standard; Protein; 118 AA.  
XX  
AC AAO01695;  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX Human polypeptide SEQ ID NO 15587.  
DE  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.

XX WO200164835-A2.  
PN 07-SEP-2001.  
XX 26-FEB-2001; 2001WO-US04927.  
PD 28-FEB-2000; 2000US-0515126.  
PF 18-MAY-2000; 2000US-0577409.  
XX (HYSE-) HYSEQ INC.  
PA Tang YT, Liu C, Drmanac RT;  
XX WPI; 2001-514838/56.  
PI N-PSDB; AAI81626.  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders.  
XX Claim 20; SEQ ID NO 15587; 1399pp + Sequence Listing; English.  
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 118 AA;  
  
Query Match  
Best Local Similarity 7.2%; Score 7; DB 22; Length 118;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 28 KKKRERK 34  
Db 31 KKKRERK 37  
  
RESULT 42  
AAG27487  
ID AAG27487 standard; Protein; 120 AA.  
XX AAG27487;  
AC AAG27487;  
XX 17-OCT-2000 (first entry)  
DT Arabidopsis thaliana protein fragment SEQ ID NO: 32343.  
DE Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
KW Arabidopsis thaliana.  
XX EPI033405-A2.  
PN 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
PF 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
PR

PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 99US-0143624.



PA (HUMA-) HUMAN GENOME SCI INC.

PI Birse CE, Rosen CA;

XX WPI; 2002-147878/19.

DR N-PSDB; ABQ56081.

XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,

PT useful in the prevention, treatment and diagnosis of cancer (e.g.

PT ovarian cancer), immune disorders, cardiovascular disorders and

PT neurological diseases -

XX Claim 11; SEQ ID NO 4136; 2922pp; English.

PS The invention relates to 2175 novel human ovarian antigens (ABP41054-

XX ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also

CC encompasses polypeptides 90% identical and polynucleotides 95% identical

CC to the sequences of the invention. The invention additionally relates to

CC recombinant vectors and host cells comprising human ovarian antigen

CC polynucleotides, antibodies against human ovarian antigen

CC of ovarian antigen polynucleotides and polypeptides in diagnosing,

CC treating, prognosing or preventing various ovary and/or breast-related

CC disorders. Such conditions include ovarian cancer and breast cancer, and

CC metastatic tumours of ovarian or breast origin, reproductive system

CC disorders (e.g., infertility, disorders of pregnancy, anovulation,

CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine

CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic

CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and

CC vaginitis), immune disorders (e.g., congenital and acquired

CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),

CC blood-related disorders (e.g., anaemia), cardiovascular disorders,

CC respiratory disorders, neurological disorders, gastrointestinal disorders,

CC and urinary system disorders. Ovarian antigen polypeptides and

CC polynucleotides may also be used in screening for compounds which

CC modulate ovarian antigen expression or activity. The polynucleotides may

CC further be used for gene therapy, chromosome mapping, in the

CC identification of individuals and in forensic analysis, and the

CC polypeptides may be used as food additives or to prepare antibodies

CC useful in disease diagnosis, drug targeting and phenotyping. The present

CC sequence represents a human ovarian antigen of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX at ftp.wipo.int/pub/published\_pct\_sequences.



us-09-854-133-586.olg.rag

Tue May 13 12:12:42 2003

PF	25-FEB-2000; 2000EP-0301439.	
XX		
PR	25-FEB-1999; 99US-0121825.	99US-0142920.
PR	05-MAR-1999; 99US-0123180.	99US-0142977.
PR	09-MAR-1999; 99US-0123548.	99US-0143542.
PR	23-MAR-1999; 99US-0125788.	99US-0143624.
PR	25-MAR-1999; 99US-0126264.	99US-0144005.
PR	29-MAR-1999; 99US-0126785.	99US-0144085.
PR	01-APR-1999; 99US-0127462.	99US-0144086.
PR	06-APR-1999; 99US-0128234.	99US-0144325.
PR	16-APR-1999; 99US-0128714.	99US-0144331.
PR	19-APR-1999; 99US-0129845.	99US-0144332.
PR	21-APR-1999; 99US-0130077.	99US-0144333.
PR	23-APR-1999; 99US-0130510.	99US-0144334.
PR	28-APR-1999; 99US-0130891.	99US-0144335.
PR	30-APR-1999; 99US-0131449.	99US-0144352.
PR	04-MAY-1999; 99US-0132048.	99US-0144884.
PR	05-MAY-1999; 99US-0132484.	99US-0144814.
PR	06-MAY-1999; 99US-0132486.	99US-0145086.
PR	07-MAY-1999; 99US-0132487.	99US-0145088.
PR	11-MAY-1999; 99US-0132863.	99US-0145085.
PR	14-MAY-1999; 99US-0134256.	99US-0145087.
PR	14-MAY-1999; 99US-0134218.	99US-0145089.
PR	14-MAY-1999; 99US-0134219.	99US-0145192.
PR	14-MAY-1999; 99US-0134221.	99US-0145145.
PR	14-MAY-1999; 99US-0134370.	99US-0145218.
PR	18-MAY-1999; 99US-0134768.	99US-0145276.
PR	19-MAY-1999; 99US-0134941.	99US-0145913.
PR	20-MAY-1999; 99US-0135124.	99US-0145918.
PR	21-MAY-1999; 99US-0135353.	99US-0145951.
PR	24-MAY-1999; 99US-0136021.	99US-0146386.
PR	25-MAY-1999; 99US-0136392.	99US-0146388.
PR	27-MAY-1999; 99US-0136782.	99US-0147038.
PR	01-JUN-1999; 99US-0137222.	99US-0147204.
PR	03-JUN-1999; 99US-0137528.	99US-0147302.
PR	04-JUN-1999; 99US-0137502.	99US-0147260.
PR	07-JUN-1999; 99US-0137724.	99US-0147303.
PR	08-JUN-1999; 99US-0138094.	99US-0147416.
PR	10-JUN-1999; 99US-0138540.	99US-0147935.
PR	14-JUN-1999; 99US-0138847.	99US-0148171.
PR	16-JUN-1999; 99US-0139119.	99US-0148319.
PR	17-JUN-1999; 99US-0139453.	99US-0148341.
PR	18-JUN-1999; 99US-0139492.	99US-0148565.
PR	18-JUN-1999; 99US-0139454.	99US-0148684.
PR	18-JUN-1999; 99US-0139455.	99US-0149368.
PR	18-JUN-1999; 99US-0139456.	99US-0149175.
PR	18-JUN-1999; 99US-0139457.	99US-0149426.
PR	18-JUN-1999; 99US-0139458.	99US-0149722.
PR	18-JUN-1999; 99US-0139459.	99US-0149723.
PR	18-JUN-1999; 99US-0139460.	99US-0149902.
PR	18-JUN-1999; 99US-0139461.	99US-0149929.
PR	18-JUN-1999; 99US-0139462.	99US-0149930.
PR	18-JUN-1999; 99US-0139463.	99US-0150566.
PR	18-JUN-1999; 99US-0139750.	99US-0150884.
PR	18-JUN-1999; 99US-0139763.	99US-0151065.
PR	21-JUN-1999; 99US-0139817.	99US-0151066.
PR	22-JUN-1999; 99US-0139899.	99US-0151080.
PR	23-JUN-1999; 99US-0140353.	99US-0151303.
PR	23-JUN-1999; 99US-0140354.	99US-0151438.
PR	24-JUN-1999; 99US-0140695.	99US-0151930.
PR	28-JUN-1999; 99US-0140823.	99US-0152363.
PR	29-JUN-1999; 99US-0140991.	99US-0153070.
PR	30-JUN-1999; 99US-0141287.	99US-0153758.
PR	01-JUL-1999; 99US-0141842.	99US-0154018.
PR	01-JUL-1999; 99US-0142154.	99US-0154039.
PR	02-JUL-1999; 99US-0142055.	99US-0154779.
PR	06-JUL-1999; 99US-0142390.	99US-0155139.
PR	08-JUL-1999; 99US-0142803.	99US-0155486.
PR		99US-0155659.
PR	09-JUL-1999;	
PR	12-JUL-1999;	
PR	13-JUL-1999;	
PR	14-JUL-1999;	
PR	15-JUL-1999;	
PR	16-JUL-1999;	
PR	16-JUL-1999;	
PR	19-JUL-1999;	
PR	19-JUL-1999;	
PR	19-JUL-1999;	
PR	19-JUL-1999;	
PR	19-JUL-1999;	
PR	19-JUL-1999;	
PR	20-JUL-1999;	
PR	20-JUL-1999;	
PR	20-JUL-1999;	
PR	21-JUL-1999;	
PR	21-JUL-1999;	
PR	21-JUL-1999;	
PR	22-JUL-1999;	
PR	22-JUL-1999;	
PR	22-JUL-1999;	
PR	22-JUL-1999;	
PR	23-JUL-1999;	
PR	23-JUL-1999;	
PR	23-JUL-1999;	
PR	26-JUL-1999;	
PR	27-JUL-1999;	
PR	27-JUL-1999;	
PR	27-JUL-1999;	
PR	28-JUL-1999;	
PR	02-AUG-1999;	
PR	02-AUG-1999;	
PR	03-AUG-1999;	
PR	04-AUG-1999;	
PR	04-AUG-1999;	
PR	05-AUG-1999;	
PR	05-AUG-1999;	
PR	06-AUG-1999;	
PR	06-AUG-1999;	
PR	09-AUG-1999;	
PR	10-AUG-1999;	
PR	11-AUG-1999;	
PR	12-AUG-1999;	
PR	13-AUG-1999;	
PR	13-AUG-1999;	
PR	16-AUG-1999;	
PR	17-AUG-1999;	
PR	18-AUG-1999;	
PR	20-AUG-1999;	
PR	20-AUG-1999;	
PR	20-AUG-1999;	
PR	23-AUG-1999;	
PR	23-AUG-1999;	
PR	25-AUG-1999;	
PR	26-AUG-1999;	
PR	27-AUG-1999;	
PR	27-AUG-1999;	
PR	30-AUG-1999;	
PR	31-AUG-1999;	
PR	01-SEP-1999;	
PR	07-SEP-1999;	
PR	10-SEP-1999;	
PR	13-SEP-1999;	
PR	15-SEP-1999;	
PR	16-SEP-1999;	
PR	20-SEP-1999;	
PR	22-SEP-1999;	
PR	23-SEP-1999;	
PR	24-SEP-1999;	

PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 21-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match

Best Local Similarity 7.2%; Score 7; DB 21; Length 132;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 KKKRERK 34

|||||||

Db 107 KKKRERK 113

Search completed: May 11, 2003, 20:10:48  
Job time : 58.6549 secs

Tue May 13 12:12:42 2003

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:09:32 ; Search time 24.0354 Seconds  
(without alignments)  
118.743 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 97  
Sequence: 1 EVEVSRDHASLGDSSETLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	6.2	7	4	US-09-258-754-407
2	6	6.2	7	4	US-09-042-107-407
3	6	6.2	27	6	5171680-14
4	6	6.2	73	4	US-09-250-609-19
5	6	6.2	88	4	US-09-134-001C-2855
6	6	6.2	113	1	US-08-168-091A-26
7	6	6.2	163	4	US-09-562-737-77
8	6	6.2	184	2	US-08-715-204-5
9	6	6.2	184	2	US-08-691-814B-50
10	6	6.2	184	3	US-09-162-597-5
11	6	6.2	184	4	US-09-250-609-13
12	6	6.2	185	2	US-08-691-814B-12
13	6	6.2	185	4	US-09-250-609-11
14	6	6.2	246	3	US-08-675-885-2
15	6	6.2	252	4	US-09-134-001C-3847
16	6	6.2	319	1	US-08-702-344-28
17	6	6.2	331	2	US-08-646-981-17
18	6	6.2	348	4	US-08-853-948B-5
19	6	6.2	350	4	US-09-651-200-25
20	6	6.2	399	1	US-08-414-926A-5
21	6	6.2	399	2	US-08-926-922-5
22	6	6.2	399	3	US-09-253-682-5
23	6	6.2	399	4	US-09-527-657-5
24	6	6.2	439	4	US-09-134-001C-4903
25	6	6.2	503	4	US-09-562-737-70
26	6	6.2	540	2	US-08-724-394A-4
27	6	6.2	635	4	US-08-931-608A-5

28	6	6.2	641	4	US-09-071-035-456	Sequence 456, App
29	6	6.2	775	2	US-08-714-070A-1	Sequence 1, Appli
30	6	6.2	1313	4	US-09-071-035-450	Sequence 450, App
31	6	6.2	1313	4	US-09-071-035-454	Sequence 454, App
32	6	6.2	1447	4	US-09-041-886-25	Sequence 25, Appl
33	6	6.2	1447	5	PCT-US94-05277-2	Sequence 2, Appli
34	6	6.2	2629	2	US-08-751-189-4	Sequence 4, Appli
35	6	6.2	2629	2	US-09-060-836-4	Sequence 4, Appli
36	6	6.2	2629	4	US-09-184-445-4	Sequence 4, Appli
37	6	6.2	3816	4	US-09-428-517-3	Sequence 3, Appli
38	6	6.2	3816	4	US-08-439-747A-2	Sequence 2, Appli
39	5	5.2	6	1	US-08-439-747A-2	Sequence 2, Appli
40	5	5.2	6	2	US-08-440-409B-2	Sequence 105, App
41	5	5.2	6	4	US-09-025-596-105	Sequence 1, Appli
42	5	5.2	8	2	US-08-701-124-1	Sequence 1, Appli
43	5	5.2	8	3	US-09-130-225-1	Sequence 1, Appli
44	5	5.2	8	4	US-09-455-061-1	Sequence 74, Appl
45	5	5.2	8	4	US-09-101-751A-74	Sequence 94, Appl
46	5	5.2	9	4	US-09-101-751A-94	Sequence 23, Appl
47	5	5.2	9	1	US-08-439-747A-23	Sequence 23, Appl
48	5	5.2	9	2	US-08-440-409B-23	Sequence 281, App
49	5	5.2	9	3	US-08-159-339A-281	Sequence 291, App
50	5	5.2	10	3	US-08-159-339A-291	Sequence 19, Appl
51	5	5.2	11	2	US-08-701-124-19	Sequence 6, Appli
52	5	5.2	11	2	US-08-623-690-6	Sequence 19, Appl
53	5	5.2	11	3	US-09-130-225-19	Sequence 19, Appl
54	5	5.2	11	4	US-09-455-061-19	Sequence 76, Appl
55	5	5.2	11	4	US-09-101-751A-76	Sequence No. 5200183
56	5	5.2	11	6	5200183-7	Patent No. 5200183
57	5	5.2	11	6	5200183-10	Sequence 16, Appl
58	5	5.2	12	1	US-08-477-509B-16	Sequence 7, Appli
59	5	5.2	12	2	US-08-623-690-7	Sequence 16, Appl
60	5	5.2	12	3	US-08-482-085B-16	Sequence 16, Appl
61	5	5.2	12	4	US-09-444-791A-16	Sequence 30, Appl
62	5	5.2	13	1	US-08-787-547-30	Sequence 8, Appli
63	5	5.2	13	2	US-08-623-690-8	Sequence No. 5168050
64	5	5.2	13	6	5168050-2	Sequence 7, Appli
65	5	5.2	15	1	US-07-720-189-7	Sequence 8, Appli
			15	1	US-07-720-189-8	

ALIGNMENTS

RESULT 1  
US-09-258-754-407  
; Sequence 407, Application US/09258754  
; Patent No. 6174687  
; GENERAL INFORMATION:  
; APPLICANT: Ruoslahti, Erkki  
; APPLICANT: Pasqualini, Renata  
; APPLICANT: Rajotte, Daniel  
; TITLE OF INVENTION: Methods of Identifying Lung Homing Molecules Using  
; TITLE OF INVENTION: Membrane Dipeptidase  
; FILE REFERENCE: P-LJ 3443  
; CURRENT APPLICATION NUMBER: US/09/258,754  
; CURRENT FILING DATE: 1999-02-26  
; EARLIER APPLICATION NUMBER: 09/042,107  
; EARLIER FILING DATE: 1998-03-13  
; NUMBER OF SEQ ID NOS: 452  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 407  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-258-754-407

Query Match 6.2%; Score 6; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 LTGGCL 83

Db 2 LTGGCL 7  
|||||

## RESULT 2

US-09-042-107-407  
; Sequence 407, Application US/09042107  
; Patent No. 623287  
; GENERAL INFORMATION:  
; APPLICANT: Ruoslahti, Erkki  
; APPLICANT: Pasqualini, Renata  
; TITLE OF INVENTION: Molecules that Home to Various Selected Organs or  
; TITLE OF INVENTION: Tissues  
; FILE REFERENCE: P-LJ 2892  
; CURRENT APPLICATION NUMBER: US/09/042,107  
; CURRENT FILING DATE: 1998-03-13  
; NUMBER OF SEQ ID NOS: 436  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 407  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-042-107-407

## Query Match

Best Local Similarity 6.2%; Score 6; DB 4; Length 7;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 LTGGCL 83  
|||||

Db 2 LTGGCL 7

## RESULT 3

5171680-14  
; Patent No. 5171680  
; APPLICANT: MULLENBACH, GUY T.; HALLEWELL, ROBERT A.; VALEZUELA,  
; PABLO  
; TITLE OF INVENTION: SUPEROXIDE DISMUTASE ANALOGS HAVING NOVEL  
; BINDING PROPERTIES  
; NUMBER OF SEQUENCES: 15  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/561,442  
; FILING DATE: 01-AUG-1990  
; SEQ ID NO:14:  
; LENGTH: 27  
5171680-14

## Query Match

Best Local Similarity 6.2%; Score 6; DB 6; Length 27;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ERKKR 31  
|||||

Db 13 ERKKR 18

## RESULT 4

US-09-250-609-19  
; Sequence 19, Application US/09250609A  
; Patent No. 6458943  
; GENERAL INFORMATION:  
; APPLICANT: Byrne, Jennifer A.  
; TITLE OF INVENTION: Members of the D52 Gene Family  
; FILE REFERENCE: 1383.0210002  
; CURRENT APPLICATION NUMBER: US/09/250,609A  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 73

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-250-609-19

Query Match 6.2%; Score 6; DB 4; Length 73;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19  
|||||

Db 16 SETLSQ 21

## RESULT 5

US-09-134-001C-2855  
; Sequence 2855, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2855  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2855

## Query Match

Best Local Similarity 6.2%; Score 6; DB 4; Length 88;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKRK 28  
|||||

Db 43 RKKRK 48

## RESULT 6

US-08-168-091A-26  
; Sequence 26, Application US/08168091A  
; Patent No. 5665862  
; GENERAL INFORMATION:  
; APPLICANT: Fischbach, Gerald.  
; APPLICANT: Falls, Douglas R.  
; APPLICANT: Rosen, Kenneth M.  
; APPLICANT: Corfas, Gabriel  
; TITLE OF INVENTION: Neurotrophic Factor  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE AND COCKFIELD  
; STREET: 60 State Street, Suite 510  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/168,091A  
; FILING DATE: 15-DEC-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/953,742



us-09-854-133-586.olig.ra

Tue May 13 12:12:42 2003

CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/715,204  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0126 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 184 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 790225  
US-08-715-204-5

Query Match 6.2%; Score 6; DB 2; Length 184;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19  
Db 100 SETLSQ 105

RESULT 9  
US-08-691-814B-50  
Sequence 50, Application US/08691814B  
Patent No. 5981218  
GENERAL INFORMATION:  
APPLICANT: Rio, Marie-Christine  
APPLICANT: Tomasetto, Catherine  
APPLICANT: Basset, Paul  
APPLICANT: Byrne, Jennifer  
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful  
TITLE OF INVENTION: as Leukemia Markers and in Breast Cancer Prognosis  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Ave, NW, Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/691,814B  
FILING DATE: 31-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/002,183

FILING DATE: 29-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: DeConti, Giulio A  
REGISTRATION NUMBER: 31,503  
REFERENCE/DOCKET NUMBER: HMI-002CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 113 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-168-091A-26

Query Match 6.2%; Score 6; DB 1; Length 113;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 29  
Db 92 KKERKK 97

RESULT 7  
US-09-562-737-77  
Sequence 77, Application US/09562737  
Patent No. 6428967  
GENERAL INFORMATION:  
APPLICANT: Herz, Joachim  
APPLICANT: Gotthardt, Michael  
TITLE OF INVENTION: LDL Receptor Signaling Pathways  
FILE REFERENCE: UTSW0708  
CURRENT APPLICATION NUMBER: US/09/562,737  
CURRENT FILING DATE: 2000-05-01  
NUMBER OF SEQ ID NOS: 132  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 77  
LENGTH: 163  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
OTHER INFORMATION: Sequence  
US-09-562-737-77

Query Match 6.2%; Score 6; DB 4; Length 163;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ASLGDS 14  
Db 13 ASLGDS 18

RESULT 8  
US-08-715-204-5  
Sequence 5, Application US/08715204  
Patent No. 5874286  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Au-Young, Janice  
APPLICANT: Goli, Surya K.  
APPLICANT: Hillman, Jennifer.  
APPLICANT: Zweiger, Gary B.  
TITLE OF INVENTION: A NOVEL TUMOR PROTEIN  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive

;; FILING DATE: 09-AUG-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Steffe, Eric K.  
;; REGISTRATION NUMBER: 36,688  
;; REFERENCE/DOCKET NUMBER: 1383.0090001  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-371-2600  
;; TELEFAX: 202-371-2543  
;; INFORMATION FOR SEQ ID NO: 50:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 184 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: not relevant  
;; TOPOLOGY: not relevant  
;; MOLECULE TYPE: peptide  
US-08-691-814B-50

Query Match  
Best Local Similarity 6.2%; Score 6; DB 2; Length 184;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 14 SETLSQ 19  
Db 100 SETLSQ 105

RESULT 10  
US-09-162-597-5  
; Sequence 5, Application US/09162597  
; Patent No. 6043343  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Hillman, Jennifer.  
; APPLICANT: Zweiger, Gary B.  
; TITLE OF INVENTION: A NOVEL TUMOR PROTEIN  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/162,597  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/715,204  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0126 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 184 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 790225

US-09-162-597-5  
Query Match  
Best Local Similarity 6.2%; Score 6; DB 3; Length 184;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 14 SETLSQ 19  
Db 100 SETLSQ 105

RESULT 11  
US-09-250-609-13  
; Sequence 13, Application US/09250609A  
; Patent No. 6458943  
; GENERAL INFORMATION:  
; APPLICANT: Byrne, Jennifer A.  
; TITLE OF INVENTION: Members of the D52 Gene Family  
; FILE REFERENCE: 1383.0210002  
; CURRENT APPLICATION NUMBER: US/09/250,609A  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 184  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-250-609-13

Query Match  
Best Local Similarity 6.2%; Score 6; DB 4; Length 184;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 14 SETLSQ 19  
Db 100 SETLSQ 105

RESULT 12  
US-08-691-814B-12  
; Sequence 12, Application US/08691814B  
; Patent No. 5981218  
; GENERAL INFORMATION:  
; APPLICANT: Rio, Marie-Christine  
; APPLICANT: Tomasetto, Catherine  
; APPLICANT: Basset, Paul  
; APPLICANT: Byrne, Jennifer  
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful  
; as Leukemia Markers and in Breast Cancer Prognosis  
; NUMBER OF SEQUENCES: 124  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Ave, NW, Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/691,814B  
; FILING DATE: 31-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/002,183  
; FILING DATE: 09-AUG-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Steffe, Eric K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1383.0090001

us-09-854-133-586.olig.ra1

Tue May 13 12:12:42 2003

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-371-2600  
 TELEFAX: 202-371-2543  
 INFORMATION FOR SEQ ID NO: 12:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 185 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-691-814B-12

Query Match 6.2%; Score 6; DB 2; Length 185;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19  
 Db 100 SETLSQ 105

RESULT 13  
 US-09-250-609-11  
 ; Sequence 11, Application US/09250609A  
 ; Patent No. 6458943  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Byrne, Jennifer A.  
 ; TITLE OF INVENTION: Members of the D52 Gene Family  
 ; FILE REFERENCE: 1383.0210002  
 ; CURRENT APPLICATION NUMBER: US/09/250.609A  
 ; CURRENT FILING DATE: 1999-02-17  
 ; NUMBER OF SEQ ID NOS: 108  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 11  
 ; LENGTH: 185  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-250-609-11

Query Match 6.2%; Score 6; DB 4; Length 185;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19  
 Db 100 SETLSQ 105

RESULT 14  
 US-08-675-885-2  
 ; Sequence 2, Application US/08675885  
 ; Patent No. 6066723  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Grammatikakis, Nicholas  
 ; APPLICANT: Grammatikakis, Aliki  
 ; APPLICANT: Toole, Bryan P.  
 ; APPLICANT: Cochran, Brent  
 ; TITLE OF INVENTION: NUCLEIC ACID ENCODING VERTEBRATE CDC37  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kevin M. Farrell, P.C.  
 ; STREET: P.O. Box 999  
 ; CITY: York Harbor  
 ; STATE: ME  
 ; COUNTRY: US  
 ; ZIP: 03909  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/675,885  
 ; FILING DATE:

CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Farrell, Kevin M.  
 REGISTRATION NUMBER: 35,505  
 REFERENCE/DOCKET NUMBER: TU-9601  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 207 3630558  
 TELEFAX: 207 3630528  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 246 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-675-885-2

Query Match 6.2%; Score 6; DB 3; Length 246;  
 Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 ELRKE 26  
 Db 3 ELRKE 8

RESULT 15  
 US-09-134-001C-3847  
 ; Sequence 3847, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 3847  
 ; LENGTH: 252  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-3847

Query Match 6.2%; Score 6; DB 4; Length 252;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 KKKRR 33  
 Db 25 KKKRR 30

RESULT 16  
 US-08-702-344-28  
 ; Sequence 28, Application US/08702344  
 ; Patent No. 5723315  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jacobs, Kenneth  
 ; APPLICANT: McCoy, John  
 ; APPLICANT: Lavallie, Edward  
 ; APPLICANT: Racie, Lisa  
 ; APPLICANT: Merberg, David  
 ; APPLICANT: Treacy, Maurice  
 ; APPLICANT: Spaulding, Vikki  
 ; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
 ; NUMBER OF SEQUENCES: 37  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genetics Institute, Inc.

;; STREET: 87 CambridgePark Drive  
;; CITY: Cambridge  
;; STATE: Massachusetts  
;; COUNTRY: U.S.A.  
;; ZIP: 02140  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/702,344  
;; FILING DATE:  
;; CLASSIFICATION: 536  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Brown, Scott A.  
;; REGISTRATION NUMBER: 32,724  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 498-8224  
;; TELEFAX: (617) 876-5851  
;; INFORMATION FOR SEQ ID NO: 28:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 319 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-702-344-28

Query Match 6.2%; Score 6; DB 1; Length 319;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TLSQTE 21  
Db 247 TLSQTE 252

RESULT 17  
US-08-646-981-17  
; Sequence 17, Application US/08646981  
; Patent No. 5852183  
; GENERAL INFORMATION:  
; APPLICANT: MAEDA, HIROAKI  
; APPLICANT: EDA, YASUYUKI  
; APPLICANT: KIMACHI, KAZUHIKO  
; APPLICANT: ONO, YOICHI  
; APPLICANT: TOKIYOSHI, SACHIO  
; TITLE OF INVENTION: DOG-MOUSE HETEROHYBRIDOMA AND GENE  
; TITLE OF INVENTION: FRAGMENT CODING FOR CONSTANT REGION OF CANINE  
; TITLE OF INVENTION: IMMUNOGLOBULINS  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/646,981  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WEINER, MARC S  
; REGISTRATION NUMBER: 32,181  
; REFERENCE/DOCKET NUMBER: 1488-106  
; INFORMATION FOR SEQ ID NO: 17:

;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 331 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
US-08-646-981-17

Query Match 6.2%; Score 6; DB 2; Length 331;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 PPSPE 71  
Db 233 PPSPE 238

RESULT 18  
US-08-853-948B-5  
; Sequence 5, Application US/08853948B  
; Patent No. 6210943  
; GENERAL INFORMATION:  
; APPLICANT: AKIHAMA, Toyota  
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE FROM CITRUS AND DNA ENCODING  
; TITLE OF INVENTION: THE SAME  
; FILE REFERENCE: 0049-0235-0  
; CURRENT APPLICATION NUMBER: US/08/853,948B  
; CURRENT FILING DATE: 1997-05-09  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 348  
; TYPE: PRT  
; ORGANISM: Citrus unshiu  
US-08-853-948B-5

Query Match 6.2%; Score 6; DB 4; Length 348;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 EMLTGG 81  
Db 42 EMLTGG 47

RESULT 19  
US-09-651-200-25  
; Sequence 25, Application US/09651200  
; Patent No. 6429303  
; GENERAL INFORMATION:  
; APPLICANT: Green et al  
; TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B  
; TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and  
; TITLE OF INVENTION: Polypeptides Encoded Thereby  
; FILE REFERENCE: 15966-562 (CURA-62)  
; CURRENT APPLICATION NUMBER: US/09/651,200  
; CURRENT FILING DATE: 2000-08-30  
; PRIOR APPLICATION NUMBER: 60/152383  
; PRIOR FILING DATE: 1999-09-03  
; PRIOR APPLICATION NUMBER: 60/172909  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 60/183578  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 25  
; LENGTH: 350  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-651-200-25

Query Match 6.2%; Score 6; DB 4; Length 350;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;



Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKKRE 32  
|||||  
Db 280 RKKKRE 285

## RESULT 20

US-08-414-926A-5  
; Sequence 5, Application US/08414926A  
; Patent No. 5721354  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; APPLICANT: Cha, Tai-An  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum  
; STREET: 5 Palo Alto Square  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-2155  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/414,926A  
; FILING DATE: March 31, 1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: AVIR-011/OOUS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-494-7622  
; TELEFAX: 415-857-0663  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 399 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-414-926A-5

Query Match 6.2%; Score 6; DB 1; Length 399;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 FIIFWI 48  
|||||  
Db 20 FIIFWI 25

## RESULT 21

US-08-926-922-5  
; Sequence 5, Application US/08926922  
; Patent No. 5925751  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; APPLICANT: Cha, Tai-An  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Luann Cserr Attorney at Law  
; STREET: 750 Arimo Avenue  
; CITY: Oakland  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94610  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/926,922  
; FILING DATE: September 10, 1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: AVIR 11A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-834-1448  
; TELEFAX: 510-839-7810  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 399 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-926-922-5

Query Match 6.2%; Score 6; DB 2; Length 399;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 FIIFWI 48  
|||||  
Db 20 FIIFWI 25

## RESULT 22

US-09-253-682-5  
; Sequence 5, Application US/09253682  
; Patent No. 6040170  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; APPLICANT: Cha, Tai-An  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Luann Cserr Attorney at Law  
; STREET: 750 Arimo Avenue  
; CITY: Oakland  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94610  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/253,682  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/926,922  
; FILING DATE: September 10, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: AVIR 11A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-834-1448  
; TELEFAX: 510-839-7810  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 399 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein.

US-09-253-682-5

Query Match 6.2%; Score 6; DB 3; Length 399;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 FIFWI 48  
| | | | |  
Db 20 FIFWI 25

RESULT 23

US-09-527-657-5  
; Sequence 5, Application US/09527657  
; Patent No. 6291236  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; Cha, Tai-An  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Luann Cserr Attorney at Law  
; STREET: 750 Arimo Avenue  
; CITY: Oakland  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94610

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/527,657  
FILING DATE: 17-Mar-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/926,922  
FILING DATE: September 10, 1997

ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: AVIR 11A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 510-834-1448

TELEFAX: 510-839-7810

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 399 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 5;

US-09-527-657-5

Query Match

Best Local Similarity 6.2%; Score 6; DB 4; Length 399;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 FIFWI 48  
| | | | |  
Db 20 FIFWI 25

RESULT 24

US-09-134-001C-4903  
; Sequence 4903, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4903  
; LENGTH: 439  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4903

Query Match 6.2%; Score 6; DB 4; Length 439;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TELRKC 25  
| | | | |  
Db 31 TELRKC 36

RESULT 25

US-09-562-737-70  
; Sequence 70, Application US/09562737  
; Patent No. 6428967  
; GENERAL INFORMATION:

APPLICANT: Herz, Joachim

APPLICANT: Gotthardt, Michael

TITLE OF INVENTION: LDL Receptor Signaling Pathways

FILE REFERENCE: UTSW0708

CURRENT APPLICATION NUMBER: US/09/562,737

CURRENT FILING DATE: 2000-05-01

NUMBER OF SEQ ID NOS: 132

SOFTWARE: PatentIn ver. 2.1

SEQ ID NO 70

LENGTH: 503

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic

OTHER INFORMATION: Sequence

US-09-562-737-70

Query Match

Best Local Similarity 6.2%; Score 6; DB 4; Length 503;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 LGRRKC 96  
| | | | |  
Db 416 LGRRKC 421

RESULT 26

US-08-724-394A-4  
; Sequence 4, Application US/08724394A  
; Patent No. 5872237  
; GENERAL INFORMATION:

APPLICANT: Feder, John N.

APPLICANT: Kronmal, Gregory S.

APPLICANT: Lauer, Peter M.

APPLICANT: Ruddy, David A.

APPLICANT: Thomas, Winston

APPLICANT: Tsuchihashi, Zenta

APPLICANT: Wolff, Roger K.

TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1

TITLE OF INVENTION: Sequences and Antibodies Thereto

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: CA

us-09-854-133-586.olig.ra1

Tue May 13 12:12:42 2003

SEQUENCE CHARACTERISTICS:  
 LENGTH: 635 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 US-08-931-608A-5

Query Match 6.2%; Score 6; DB 4; Length 635;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ASLGDS 14  
 Db 194 ASLGDS 199

RESULT 28  
 US-09-071-035-456  
 ; Sequence 456, Application US/09071035  
 ; Patent No. 6448043  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gil H. Choi  
 ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
 ; NUMBER OF SEQUENCES: 496  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/071,035  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: A. Anders Brooks  
 ; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PB369P2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 456:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 641 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-09-071-035-456

Query Match 6.2%; Score 6; DB 4; Length 641;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ETLST 20  
 Db 241 ETLST 246

RESULT 29  
 US-08-714-070A-1  
 ; Sequence 1, Application US/08714070A

COUNTRY: USA  
 ZIP: 94111-3834  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/724,394A  
 FILING DATE: 01-OCT-1996  
 CLASSIFICATION: 536  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fitts, Renee A.  
 REGISTRATION NUMBER: 35,136  
 REFERENCE/DOCKET NUMBER: 017957-000100  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-576-0200  
 TELEFAX: 415-576-0300  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 540 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: peptide  
 FEATURE:  
 NAME/KEY: Region  
 LOCATION: 1..540  
 OTHER INFORMATION: /note= "BTF5"

US-08-724-394A-4

Query Match 6.2%; Score 6; DB 2; Length 540;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKRE 32  
 Db 298 RKKRE 303

RESULT 27  
 US-08-931-608A-5  
 ; Sequence 5, Application US/08931608A  
 ; Patent No. 6302685  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lobel, Peter  
 ; APPLICANT: Sleat, David E.  
 ; TITLE OF INVENTION: NOVEL HUMAN LYSOSOMAL PROTEIN AND METHODS OF ITS USE  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: David A. Jackson, Esq.  
 ; STREET: 411 Hackensack Ave, Continental Plaza, 4th Floor  
 ; CITY: Hackensack  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 07601  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/931,608A  
 ; FILING DATE:  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jackson Esq., David A.  
 ; REGISTRATION NUMBER: 26,742  
 ; REFERENCE/DOCKET NUMBER: 601-1-077  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 201-487-5800  
 ; TELEFAX: 201-343-1684  
 ; INFORMATION FOR SEQ ID NO: 5:

; Patent No. 5834237  
; GENERAL INFORMATION:  
; APPLICANT: JACOBS, Eric  
; APPLICANT: SILVESTRE, Nathalie  
; APPLICANT: SCHWEINBRUBER, Ernst  
; TITLE OF INVENTION: COMBINED USE OF TWO EXPRESSION CASSETTES  
; TITLE OF INVENTION: FOR THE PRODUCTION OF A PROTEIN OF INTEREST  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/714,070A  
; FILING DATE: 07-OCT-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 94/01767  
; FILING DATE: 10-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa Stanek  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 017753-077  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 775 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Schizosaccharomyces pombe  
; US-08-714-070A-1

Query Match 6.2%; Score 6; DB 2; Length 775;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 26 ERKKR 31  
Db 20 ERKKR 25

RESULT 30  
US-09-071-035-450  
; Sequence 450, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 450:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1313 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-071-035-450

Query Match 6.2%; Score 6; DB 4; Length 1313;  
Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 15 ETLSQT 20  
Db 877 ETLSQT 882

RESULT 31  
US-09-071-035-454  
; Sequence 454, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 454:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1313 amino acids  
; TYPE: amino acid



us-09-854-133-586.olig.ra1

Tue May 13 12:12:42 2003

```

;
; APPLICANT: Zabrecky, James R.
; TITLE OF INVENTION: Antibodies Specific for DCC Gene Product
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05277
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.42709
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202.508.9100
; TELEFAX: 202.508.9299
; TELEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-05277-2
;
; Query Match 6.2%; Score 6; DB 5; Length 1447;
; Best Local Similarity 100.0%; Pred. No. 4.8e+02;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 9 ASLGDS 14
Db 119 ASLGDS 124
;
; RESULT 34
; US-08-751-189-4
; Sequence 4, Application US/08751189
; Patent No. 5919656
; GENERAL INFORMATION:
; APPLICANT: Harrington, Lea A.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 5919656el Genes Encoding Telomerase Protein
; TITLE OF INVENTION: 1
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen, Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,189
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy A.
; REGISTRATION NUMBER: 34,688

```

```

;
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-071-035-454
;
; Query Match 6.2%; Score 6; DB 4; Length 1313;
; Best Local Similarity 100.0%; Pred. No. 4.4e+02;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 15 ETLSTQ 20
Db 877 ETLSTQ 882
;
; RESULT 32
; US-09-041-886-25
; Sequence 25, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-041-886-25
;
; Query Match 6.2%; Score 6; DB 4; Length 1447;
; Best Local Similarity 100.0%; Pred. No. 4.8e+02;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 9 ASLGDS 14
Db 119 ASLGDS 124
;
; RESULT 33
; PCT-US94-05277-2
; Sequence 2, Application PC/TUS9405277
; GENERAL INFORMATION:
; APPLICANT: Bruskin, Arthur
; APPLICANT: Jarosz, David E.
; APPLICANT: Johnson, Karen
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: Vogelstein, Bert

```

```
; REFERENCE/DOCKET NUMBER: A-433
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2629 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-751-189-4
```

```
Query Match
Best Local Similarity 6.2%; Score 6; DB 2; Length 2629;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 70 KEVTCR 75
Db 1077 KEVTCR 1082
|||||
```

```
RESULT 35
US-09-060-836-4
; Sequence 4, Application US/09060836
; Patent No. 5981707
; GENERAL INFORMATION:
; APPLICANT: Harrington, Lea A.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 5981707el Genes Encoding Telomerase Protein
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen, Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,836
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/751,189
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-433
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2629 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-09-060-836-4
```

```
Query Match
Best Local Similarity 6.2%; Score 6; DB 2; Length 2629;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 70 KEVTCR 75
Db 1077 KEVTCR 1082
|||||
```

```
RESULT 36
US-09-184-445-4
; Sequence 4, Application US/09184445
```

```
; Patent No. 6174703
; GENERAL INFORMATION:
; APPLICANT: Harrington, Lea A.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 6174703el Genes Encoding Telomerase Protein
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen, Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/184,445
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/751,189
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-433
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2629 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-09-184-445-4
```

```
Query Match
Best Local Similarity 6.2%; Score 6; DB 4; Length 2629;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 70 KEVTCR 75
Db 1077 KEVTCR 1082
|||||
```

```
RESULT 37
US-09-428-517-3
; Sequence 3, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER FILING DATE: 1999-02-16
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3816
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
```



Db 1 RKKR 5

RESULT 41  
US-08-701-124-1  
; Sequence 1, Application US/08701124  
; Patent No. 5846782  
; GENERAL INFORMATION:  
; APPLICANT: Wickham, Thomas J.  
; APPLICANT: Roelvink, Petrus W.  
; APPLICANT: Kovesdi, Imre  
; TITLE OF INVENTION: TARGETING ADENOVIRUS WITH USE OF  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
; STREET: Two Prudential Plaza - 49th Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/701,124  
; FILING DATE: 21-AUG-1996  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-701-124-1

Query Match  
Best Local Similarity 5.2%; Score 5; DB 2; Length 8;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKR 31  
Db 1 RKKR 5

RESULT 42  
US-09-130-225-1  
; Sequence 1, Application US/09130225  
; Patent No. 6057155  
; GENERAL INFORMATION:  
; APPLICANT: Wickham, Thomas J.  
; APPLICANT: Roelvink, Petrus W.  
; APPLICANT: Kovesdi, Imre  
; TITLE OF INVENTION: TARGETING ADENOVIRUS WITH USE OF  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
; STREET: Two Prudential Plaza - 49th Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/130,225  
; FILING DATE:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 8-701124  
; FILING DATE: 21-AUG-1996  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-130-225-1

Query Match  
Best Local Similarity 5.2%; Score 5; DB 3; Length 8;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKR 31  
Db 1 RKKR 5

RESULT 43  
US-09-455-061-1  
; Sequence 1, Application US/09455061  
; Patent No. 6329190  
; GENERAL INFORMATION:  
; APPLICANT: Wickham, Thomas J.  
; APPLICANT: Roelvink, Petrus W.  
; APPLICANT: Kovesdi, Imre  
; TITLE OF INVENTION: TARGETING ADENOVIRUS WITH USE OF  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
; STREET: Two Prudential Plaza - 49th Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/455,061  
; FILING DATE: 06-DEC-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 9-130225  
; FILING DATE: 06-AUG-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 8-701124  
; FILING DATE: 21-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hefner, M. Daniel  
; REGISTRATION NUMBER: 41,826  
; REFERENCE/DOCKET NUMBER: 203128  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-455-061-1

Query Match  
Best Local Similarity 5.2%; Score 5; DB 4; Length 8;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKR 31  
Db 1 RKKR 5



us-09-854-133-586.olig.ra1

Tue May 13 12:12:42 2003

; OTHER INFORMATION: Description of Unknown Organism: Artificial  
; OTHER INFORMATION: Sequence  
US-09-101-751A-94

Query Match 5.2%; Score 5; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKR 31  
|  
|  
|  
|  
Db 1 RKKR 5

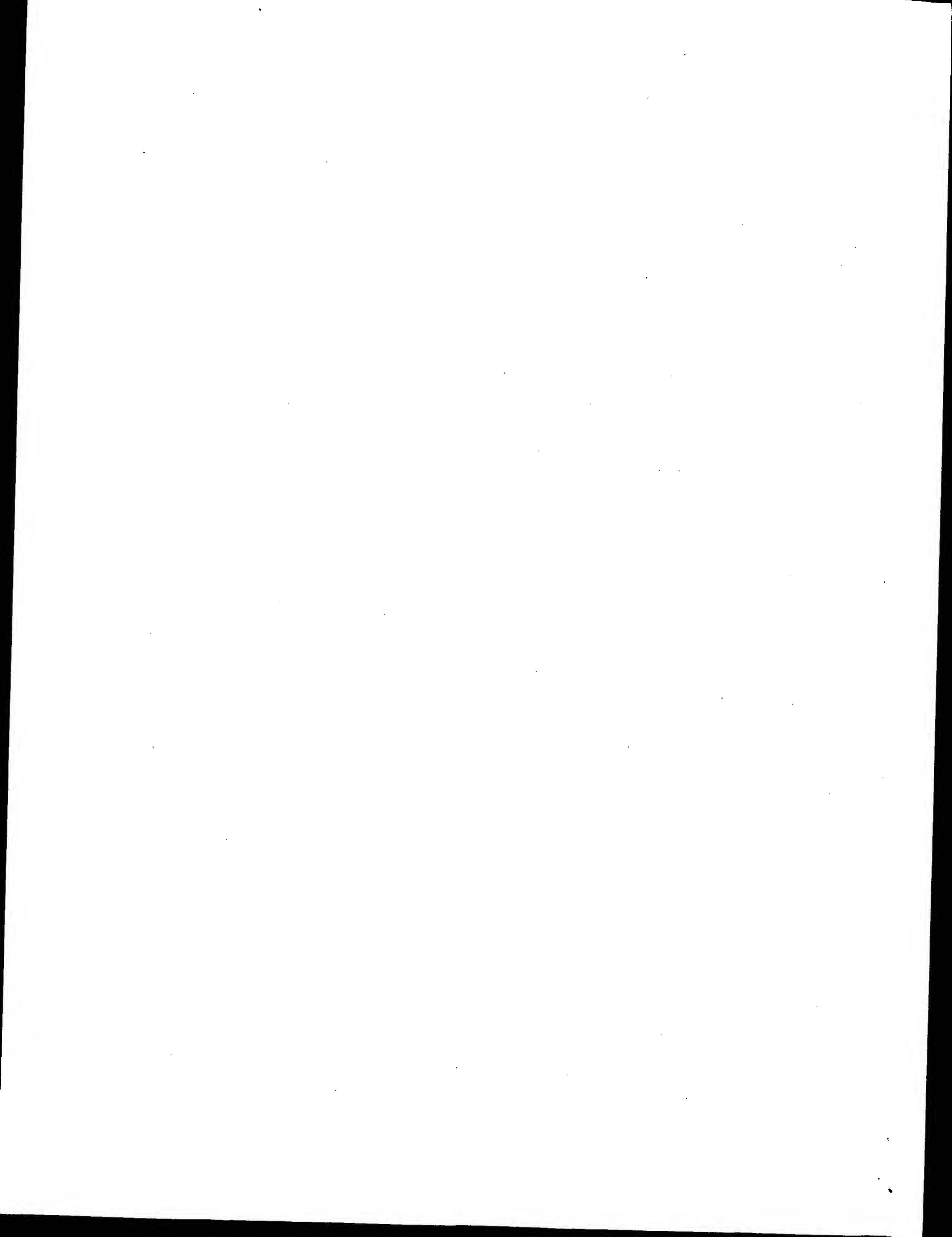
Search completed: May 11, 2003, 20:14:26  
Job time : 28.0354 secs

RESULT 44  
US-09-101-751A-74  
; Sequence 74, Application US/09101751A  
; Patent No. 6465253  
; GENERAL INFORMATION:  
; APPLICANT: WICKHAM, THOMAS J.  
; APPLICANT: KOVESDI, IMRE  
; APPLICANT: BROUGH, DOUGLAS E.  
; TITLE OF INVENTION: VECTORS AND METHODS FOR GENE TRANSFER TO CELLS  
; FILE REFERENCE: 85710  
; CURRENT APPLICATION NUMBER: US/09/101,751A  
; CURRENT FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: WO 96US19150  
; PRIOR FILING DATE: 1996-11-27  
; PRIOR APPLICATION NUMBER: US 08/700,846  
; PRIOR FILING DATE: 1996-08-21  
; PRIOR APPLICATION NUMBER: US 08/701,124  
; PRIOR FILING DATE: 1996-08-21  
; PRIOR APPLICATION NUMBER: US 08/563,368  
; PRIOR FILING DATE: 1995-11-28  
; NUMBER OF SEQ ID NOS: 94  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 74  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: ().()  
; OTHER INFORMATION: Description of Unknown Organism: Artificial  
; OTHER INFORMATION: Sequence  
US-09-101-751A-74

Query Match 5.2%; Score 5; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKR 31  
|  
|  
|  
|  
Db 1 RKKR 5

RESULT 45  
US-09-101-751A-94  
; Sequence 94, Application US/09101751A  
; Patent No. 6465253  
; GENERAL INFORMATION:  
; APPLICANT: WICKHAM, THOMAS J.  
; APPLICANT: KOVESDI, IMRE  
; APPLICANT: BROUGH, DOUGLAS E.  
; TITLE OF INVENTION: VECTORS AND METHODS FOR GENE TRANSFER TO CELLS  
; FILE REFERENCE: 85710  
; CURRENT APPLICATION NUMBER: US/09/101,751A  
; CURRENT FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: WO 96US19150  
; PRIOR FILING DATE: 1996-11-27  
; PRIOR APPLICATION NUMBER: US 08/700,846  
; PRIOR FILING DATE: 1996-08-21  
; PRIOR APPLICATION NUMBER: US 08/701,124  
; PRIOR FILING DATE: 1996-08-21  
; PRIOR APPLICATION NUMBER: US 08/563,368  
; PRIOR FILING DATE: 1995-11-28  
; NUMBER OF SEQ ID NOS: 94  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 94  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: ().()  
; OTHER INFORMATION: Sequence



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:12:57 ; Search time 31.7611 Seconds  
(without alignments)  
281.051 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 97  
Sequence: 1 EVEVSRDHASLGDSETLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size : 0

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : Published\_Applications\_AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	100.0	97	9 US-09-854-133-586	Sequence 586, App
2	97	100.0	97	10 US-09-738-973-586	Sequence 586, App
3	16	16.5	16	9 US-09-854-133-587	Sequence 587, App
4	16	16.5	16	10 US-09-738-973-587	Sequence 587, App
5	7	7.2	36	9 US-09-764-872-359	Sequence 359, App
6	7	7.2	47	9 US-10-091-504-1018	Sequence 1018, App
7	7	7.2	47	10 US-09-764-869-1018	Sequence 1018, App
8	7	7.2	66	10 US-09-764-877-1089	Sequence 1089, App
9	7	7.2	94	9 US-09-866-050A-676	Sequence 676, App
10	7	7.2	170	9 US-09-764-868-816	Sequence 816, App
11	7	7.2	517	9 US-10-280-403-6	Sequence 6, Appli
12	7	7.2	517	10 US-09-907-479-6	Sequence 6, Appli
13	7	7.2	802	10 US-09-876-889-352	Sequence 352, App
14	6	6.2	19	9 US-09-984-245-306	Sequence 306, App
15	6	6.2	19	9 US-09-966-262-306	Sequence 306, App
16	6	6.2	19	9 US-09-983-966-306	Sequence 306, App
17	6	6.2	19	9 US-10-143-090-306	Sequence 306, App
18	6	6.2	28	9 US-09-925-299-1023	Sequence 1023, App
19	6	6.2	28	10 US-09-925-299-1023	Sequence 1023, App

20	6	6.2	35	9 US-10-001-857-134	Sequence 134, App
21	6	6.2	35	12 US-10-001-870-145	Sequence 145, App
22	6	6.2	46	9 US-09-764-891-4700	Sequence 4700, App
23	6	6.2	51	10 US-09-864-761-37309	Sequence 37309, A
24	6	6.2	51	10 US-09-864-761-42719	Sequence 42719, A
25	6	6.2	59	10 US-09-864-761-43079	Sequence 43079, A
26	6	6.2	64	10 US-09-864-761-33889	Sequence 33889, A
27	6	6.2	69	10 US-09-864-761-34523	Sequence 34523, A
28	6	6.2	70	9 US-09-764-891-3256	Sequence 3256, App
29	6	6.2	73	10 US-09-250-611-19	Sequence 19, Appl
30	6	6.2	76	9 US-10-002-344A-229	Sequence 229, App
31	6	6.2	80	10 US-09-864-761-42011	Sequence 42011, A
32	6	6.2	89	9 US-10-072-349-99	Sequence 99, Appl
33	6	6.2	89	10 US-09-764-855-99	Sequence 99, Appl
34	6	6.2	97	9 US-10-091-504-1022	Sequence 1022, Ap
35	6	6.2	97	10 US-09-764-869-1022	Sequence 1022, Ap
36	6	6.2	101	9 US-09-925-299-774	Sequence 774, App
37	6	6.2	101	10 US-09-925-299-774	Sequence 774, App
38	6	6.2	123	9 US-10-043-487-333	Sequence 333, App
39	6	6.2	123	9 US-09-764-891-4621	Sequence 4621, Ap
40	6	6.2	128	10 US-09-864-761-36308	Sequence 36308, A
41	6	6.2	141	9 US-09-828-995B-29	Sequence 29, Appl
42	6	6.2	141	10 US-09-864-761-36181	Sequence 36181, A
43	6	6.2	151	10 US-09-789-561-88	Sequence 88, Appl
44	6	6.2	160	10 US-09-321-801-17	Sequence 17, Appl
45	6	6.2	163	9 US-10-211-962-77	Sequence 77, Appl
46	6	6.2	167	1 US-08-781-986A-5216	Sequence 5216, Ap
47	6	6.2	176	10 US-09-925-302-764	Sequence 764, App
48	6	6.2	184	10 US-09-250-611-13	Sequence 13, Appl
49	6	6.2	185	10 US-09-250-611-11	Sequence 11, Appl
50	6	6.2	230	10 US-09-925-300-1377	Sequence 1377, Ap
51	6	6.2	234	10 US-09-800-729-150	Sequence 150, App
52	6	6.2	246	9 US-10-157-223-5	Sequence 5, Appli
53	6	6.2	246	10 US-09-815-242-12986	Sequence 12986, A
54	6	6.2	246	10 US-09-815-242-13040	Sequence 13040, A
55	6	6.2	260	10 US-09-925-300-1470	Sequence 1470, Ap
56	6	6.2	261	9 US-09-738-626-6335	Sequence 6335, Ap
57	6	6.2	277	9 US-09-965-529-15	Sequence 15, Appl
58	6	6.2	281	9 US-10-101-464A-619	Sequence 619, App
59	6	6.2	290	10 US-09-910-174A-19	Sequence 19, Appl
60	6	6.2	318	10 US-09-947-971-2	Sequence 2, Appli
61	6	6.2	319	9 US-10-114-893-210	Sequence 210, App
62	6	6.2	319	12 US-10-025-335-1	Sequence 1, Appli
63	6	6.2	346	9 US-10-081-816-16	Sequence 16, Appl
64	6	6.2	350	10 US-09-910-174A-17	Sequence 17, Appl
65	6	6.2	352	9 US-09-828-995B-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1  
US-09-854-133-586  
; Sequence 586, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 586  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-854-133-586

```
Query Match      100.0%; Score 97; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. 7.1e-89;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVEVSRDHASLGDSETLSQTELKRRKKRKKRKFQANCGIDFIIFWIFWILLFHHWIIQ 60
   |||||
Db 1 EVEVSRDHASLGDSETLSQTELKRRKKRKKRKFQANCGIDFIIFWIFWILLFHHWIIQ 60
   |||||

QY 61 ESLCPSPPKVTCTREMLTGGCLPWATRSHLGRKCS 97
   |||||
Db 61 ESLCPSPPKVTCTREMLTGGCLPWATRSHLGRKCS 97
   |||||

RESULT 2
US-09-738-973-586
; Sequence 586, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 586
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-586
```

```
Query Match      100.0%; Score 97; DB 10; Length 97;
Best Local Similarity 100.0%; Pred. No. 7.1e-89;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVEVSRDHASLGDSETLSQTELKRRKKRKKRKFQANCGIDFIIFWIFWILLFHHWIIQ 60
   |||||
Db 1 EVEVSRDHASLGDSETLSQTELKRRKKRKKRKFQANCGIDFIIFWIFWILLFHHWIIQ 60
   |||||

QY 61 ESLCPSPPKVTCTREMLTGGCLPWATRSHLGRKCS 97
   |||||
Db 61 ESLCPSPPKVTCTREMLTGGCLPWATRSHLGRKCS 97
   |||||
```

```
RESULT 3
US-09-854-133-587
; Sequence 587, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
```

```
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 587
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-133-587
```

```
Query Match      16.5%; Score 16; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 3.2e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 35 FQANCGIDFIIFWIFW 50
   |||||
Db 1 FQANCGIDFIIFWIFW 16
   |||||
```

```
RESULT 4
US-09-738-973-587
; Sequence 587, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 587
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-587
```

```
Query Match      16.5%; Score 16; DB 10; Length 16;
Best Local Similarity 100.0%; Pred. No. 3.2e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 35 FQANCGIDFIIFWIFW 50
   |||||
Db 1 FQANCGIDFIIFWIFW 16
   |||||
```

```
RESULT 5
US-09-764-872-359
; Sequence 359, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 359
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-872-359
```



Tue May 13 12:12:43 2003

```

; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1089
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-877-1089

```

```

Query Match      7.2%; Score 7; DB 10; Length 66;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 24 KKERKKK 30
Db 35 KKERKKK 41

```

```

RESULT 9
US-09-866-050A-676
; Sequence 676, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 676
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Mouse
US-09-866-050A-676

```

```

Query Match      7.2%; Score 7; DB 9; Length 94;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 23 RKKKKK 29
Db 35 RKKKKK 41

```

```

RESULT 10
US-09-764-868-816
; Sequence 816, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 816
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE

```

```

Query Match      7.2%; Score 7; DB 9; Length 36;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 24 KKERKKK 30
Db 7 KKERKKK 13

```

```

RESULT 6
US-10-091-504-1018
; Sequence 1018, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C1
; CURRENT APPLICATION NUMBER: US/10/091,504
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2442
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1018
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-091-504-1018

```

```

Query Match      7.2%; Score 7; DB 9; Length 47;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 25 KKKKKK 31
Db 14 KKKKKK 20

```

```

RESULT 7
US-09-764-869-1018
; Sequence 1018, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1018
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-869-1018

```

```

Query Match      7.2%; Score 7; DB 10; Length 47;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 25 KKKKKK 31
Db 14 KKKKKK 20

```

```

RESULT 8
US-09-764-877-1089
; Sequence 1089, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

```

; LOCATION: (118)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-816

Query Match  
Best Local Similarity 7.2%; Score 7; DB 9; Length 170;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKKKK 29  
Db 17 RKKKKK 23

## RESULT 11

US-10-280-403-6  
; Sequence 6, Application US/10280403  
; Publication No. US20030082620A1  
; GENERAL INFORMATION:  
; APPLICANT: Astle, Jon H  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Dwivedi, Pooranima  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Molino, Gary A.  
; APPLICANT: Myerow, Susan H.  
; APPLICANT: Thiagalalingam, Arunthathi  
; APPLICANT: Catino, Theodore

; TITLE OF INVENTION: No. US20030082620A1el Human Genes and Gene Expression Products:  
; FILE REFERENCE: 1657/1015B  
; CURRENT APPLICATION NUMBER: US/10/280,403  
; CURRENT FILING DATE: 2002-10-25  
; PRIOR APPLICATION NUMBER: US 09/385,982  
; PRIOR FILING DATE: 1999-08-30  
; PRIOR APPLICATION NUMBER: US 09/328,111  
; PRIOR FILING DATE: 1999-06-08  
; PRIOR APPLICATION NUMBER: US 60/098,639  
; PRIOR FILING DATE: 1988-08-31  
; PRIOR APPLICATION NUMBER: US 60/117,393  
; PRIOR FILING DATE: 1998-01-27  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-280-403-6

Query Match  
Best Local Similarity 7.2%; Score 7; DB 9; Length 517;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KKKKKK 31  
Db 300 KKKKKK 306

## RESULT 12

US-09-907-479-6  
; Sequence 6, Application US/09907479  
; Patent No. US20020034758A1  
; GENERAL INFORMATION:  
; APPLICANT: Astle, Jon  
; APPLICANT: Burgess, Christopher  
; APPLICANT: Dwivedi, Pooranima  
; APPLICANT: Lewis, Marcia  
; APPLICANT: Molino, Gary  
; APPLICANT: Myerow, Susan  
; APPLICANT: Thiagalalingam, Arunthathi  
; APPLICANT: Catino, Theodore  
; TITLE OF INVENTION: No. US20020034758A1el Human Genes and Gene Expression Products:  
; FILE REFERENCE: 1657/1015B  
; CURRENT APPLICATION NUMBER: US/09/907,479  
; CURRENT FILING DATE: 2001-07-17  
; PRIOR APPLICATION NUMBER: US 09/385,982

; PRIOR FILING DATE: 1999-08-30  
; PRIOR APPLICATION NUMBER: US 09/328,111  
; PRIOR FILING DATE: 1999-06-08  
; PRIOR APPLICATION NUMBER: US 60/098,639  
; PRIOR FILING DATE: 1988-08-31  
; PRIOR APPLICATION NUMBER: US 60/117,393  
; PRIOR FILING DATE: 1998-01-27  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-907-479-6

Query Match  
Best Local Similarity 7.2%; Score 7; DB 10; Length 517;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KKKKKK 31  
Db 300 KKKKKK 306

## RESULT 13

US-09-876-889-352  
; Sequence 352, Application US/09876889  
; Patent No. US20020076715A1  
; GENERAL INFORMATION:  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: King, Gordon E.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN  
; TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS  
; FILE REFERENCE: 210121.466C3  
; CURRENT APPLICATION NUMBER: US/09/876,889  
; CURRENT FILING DATE: 2001-06-06  
; NUMBER OF SEQ ID NOS: 353  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 352  
; LENGTH: 802  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-876-889-352

Query Match  
Best Local Similarity 7.2%; Score 7; DB 10; Length 802;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 SQTELK 24  
Db 282 SQTELK 288

## RESULT 14

US-09-984-245-306  
; Sequence 306, Application US/09984245  
; Patent No. US20020165374A1  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: 87 Human Secreted Proteins  
; FILE REFERENCE: P2004P1  
; CURRENT APPLICATION NUMBER: US/09/984,245  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 09/154,707  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: PCT/US98/05311  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: US 60/041,277  
; PRIOR FILING DATE: 1997-03-21  
; PRIOR APPLICATION NUMBER: US 60/042,344  
; PRIOR FILING DATE: 1997-03-21

us-09-854-133-586.olg.rapb

Tue May 13 12:12:43 2003

PRIOR APPLICATION NUMBER: US 60/041,276  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/041,281  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/048,094  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,350  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,188  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,135  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/050,937  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,187  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,099  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,352  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,186  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,096  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,355  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,160  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,351  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,154  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/054,804  
PRIOR FILING DATE: 1997-08-05  
PRIOR APPLICATION NUMBER: US 60/056,370  
PRIOR FILING DATE: 1997-08-19  
PRIOR APPLICATION NUMBER: US 60/060,862  
PRIOR FILING DATE: 1997-10-02  
NUMBER OF SEQ ID NOS: 343  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 306  
LENGTH: 19  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-984-245-306

Query Match 6.2%; Score 6; DB 9; Length 19;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERK 29  
Db 10 KKERK 15

RESULT 15  
US-09-966-262-306  
; Sequence 306, Application US/09966262  
; Publication No. US20030050461A1  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: 87 Human Secreted Proteins  
; FILE REFERENCE: PZ004P1  
; CURRENT APPLICATION NUMBER: US/09/966,262  
; CURRENT FILING DATE: 2001-10-01  
; PRIOR APPLICATION NUMBER: US 09/154,707  
; PRIOR FILING DATE: 1998-09-17

PRIOR APPLICATION NUMBER: PCT/US98/05311  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: US 60/041,277  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/042,344  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/041,276  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/041,281  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/048,094  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,350  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,188  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,135  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/050,937  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,187  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,099  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,352  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,186  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,069  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,095  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,131  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,096  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,355  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,160  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,351  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,154  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/054,804  
PRIOR FILING DATE: 1997-08-05  
PRIOR APPLICATION NUMBER: US 60/056,370  
PRIOR FILING DATE: 1997-08-19  
PRIOR APPLICATION NUMBER: US 60/060,862  
PRIOR FILING DATE: 1997-10-02  
NUMBER OF SEQ ID NOS: 343  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 306  
LENGTH: 19  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-966-262-306

Query Match 6.2%; Score 6; DB 9; Length 19;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERK 29  
Db 10 KKERK 15

RESULT 16  
US-09-983-966-306  
; Sequence 306, Application US/09983966  
; Publication No. US20030060619A1  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.

```
; TITLE OF INVENTION: 87 Human Secreted Proteins
; FILE REFERENCE: P2004P1
; CURRENT APPLICATION NUMBER: US/09/983,966
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/154,707
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: PCT/US98/05311
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: US 60/041,277
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/042,344
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/041,276
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/041,281
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/048,094
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,350
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,188
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,135
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/050,937
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,187
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,099
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,352
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,186
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,069
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,095
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,131
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,096
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,355
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,160
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,351
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,154
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/054,804
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: US 60/056,370
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: US 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 343
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 306
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-983-966-306
```

```
Query Match
Best Local Similarity 6.2%; Score 6; DB 9; Length 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 29
Db 10 KKERKK 15
```

```
RESULT 17
US-10-143-090-306
; Sequence 306, Application US/10143090
; Publication No. US20030069406A1
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 87 Human Secreted Proteins
; FILE REFERENCE: P2004P1
; CURRENT APPLICATION NUMBER: US/10/143,090
; PRIOR FILING DATE: 2002-05-13
; PRIOR APPLICATION NUMBER: 09/154,707
; PRIOR FILING DATE: 1998-09-17
; NUMBER OF SEQ ID NOS: 343
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 306
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-143-090-306
```

```
Query Match
Best Local Similarity 6.2%; Score 6; DB 9; Length 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 29
Db 10 KKERKK 15
```

```
RESULT 18
US-09-925-299-1023
; Sequence 1023, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1023
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-1023
```

```
Query Match
Best Local Similarity 6.2%; Score 6; DB 9; Length 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19
Db 14 SETLSQ 19
```

```
RESULT 19
US-09-925-299-1023
; Sequence 1023, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
```



us-09-854-133-586.olig.rapb

Tue May 13 12:12:43 2003

```

; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1023
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-1023

Query Match
Best Local Similarity 100.0%; Score 6; DB 10; Length 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19
Db 14 SETLSQ 19

RESULT 20
US-10-001-857-134
; Sequence 134, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 134
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-857-134

Query Match
Best Local Similarity 100.0%; Score 6; DB 9; Length 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19
Db 20 SETLSQ 25

RESULT 21
US-10-001-870-145
; Sequence 145, Application US/10001870
; Patent No. US20020150924A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and
; FILE REFERENCE: DEX-0283
; CURRENT APPLICATION NUMBER: US/10/001,870
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,189
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 145
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapien

US-10-001-870-145

Query Match
Best Local Similarity 100.0%; Score 6; DB 12; Length 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKK 30
Db 11 KERKKK 16

RESULT 22
US-09-764-891-4700
; Sequence 4700, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4700
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (28)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4700

Query Match
Best Local Similarity 100.0%; Score 6; DB 9; Length 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19
Db 35 SETLSQ 40

RESULT 23
US-09-864-761-37309
; Sequence 37309, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30

```

;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 37309  
;; LENGTH: 51  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC010885.2  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7  
US-09-864-761-37309

Query Match 6.2%; Score 6; DB 10; Length 51;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 KKRERK 34  
|||||  
Db 4 KKRERK 9

RESULT 24  
US-09-864-761-42719  
;; Sequence 42719, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; FILE REFERENCE: Aecomica-x-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27

;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 42719  
;; LENGTH: 51  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC010885.1  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
US-09-864-761-42719

Query Match 6.2%; Score 6; DB 10; Length 51;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 KKRERK 34  
|||||  
Db 4 KKRERK 9

RESULT 25  
US-09-864-761-43079  
;; Sequence 43079, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; FILE REFERENCE: Aecomica-x-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30

us-09-854-133-586.olig.rapb

Tue May 13 12:12:43 2003

```

; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33889
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007546.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EST_HUMAN HIT: AW867799.1, EVALUE 3.00e-13
; OTHER INFORMATION: EST_HUMAN HIT: AUL25286.1, EVALUE 4.00e-13
; US-09-864-761-33889

Query Match          6.2%; Score 6; DB 10; Length 64;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 IQESLL 64
Db 2 IQESLL 7

RESULT 27
US-09-864-761-34523
; Sequence 34523, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43079
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL049735.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.68
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.79
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.65
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.63
; US-09-864-761-43079

Query Match          6.2%; Score 6; DB 10; Length 59;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKRE 32
Db 6 RKKRE 11

RESULT 26
US-09-864-761-33889
; Sequence 33889, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
```

```
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34523
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006028.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 14
; OTHER INFORMATION: EST_HUMAN HIT: BE071726.1, EVALUE 5.40e-01
US-09-864-761-34523
```

```
Query Match
Best Local Similarity 6.2%; Score 6; DB 10; Length 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 25 KERKK 30
    |||||
Db 36 KERKK 41
```

```
RESULT 28
US-09-764-891-3256
; Sequence 3256, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
```

```
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3256
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (15)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3256
```

```
Query Match
Best Local Similarity 6.2%; Score 6; DB 9; Length 70;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 28 KKKRR 33
    |||||
Db 1 KKKRR 6
```

```
RESULT 29
US-09-250-611-19
; Sequence 19, Application US/09250611
; Patent No. US20020143161A1
; GENERAL INFORMATION:
; APPLICANT: Byrne, Jennifer A.
; APPLICANT: Basset, Paul
; TITLE OF INVENTION: Members of the D52 Gene Family
; FILE REFERENCE: 1383.0210001
; CURRENT APPLICATION NUMBER: US/09/250,611
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-250-611-19
```

```
Query Match
Best Local Similarity 6.2%; Score 6; DB 10; Length 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 14 SETLSQ 19
    |||||
Db 16 SETLSQ 21
```

```
RESULT 30
US-10-002-344A-229
; Sequence 229, Application US/10002344A
; Patent No. US20020172959A1
; GENERAL INFORMATION:
; APPLICANT: Recipon, Herve
; APPLICANT: Sun, Yongming
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and
; FILE REFERENCE: DEX-0241
; CURRENT APPLICATION NUMBER: US/10/002,344A
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/242,998
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 229
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapien
```



Tue May 13 12:12:43 2003

US-10-002-344A-229  
Query Match 6.2%; Score 6; DB 9; Length 76;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 29  
Db 62 KKERKK 67

RESULT 31  
US-09-864-761-42011  
; Sequence 42011, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aomic-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 42011  
; LENGTH: 80  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC007340.3  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.6  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 7.3  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.8

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.9  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 9.8  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9.4  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 8.5  
OTHER INFORMATION: EST\_HUMAN HIT: AV718492.1, EVALUE 3.90e+00  
US-09-864-761-42011

Query Match 6.2%; Score 6; DB 10; Length 80;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 29  
Db 44 KKERKK 49

RESULT 32  
US-10-072-349-99  
; Sequence 99, Application US/10072349  
; Publication No. US20030054420A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P110C1  
; CURRENT APPLICATION NUMBER: US/10/072,349  
; CURRENT FILING DATE: 2002-02-11  
; Prior Application removed - See file Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 334  
; SOFTWARE: PatentIn Ver. 3.1  
; SEQ ID NO 99  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-072-349-99

Query Match 6.2%; Score 6; DB 9; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19  
Db 45 SETLSQ 50

RESULT 33  
US-09-764-855-99  
; Sequence 99, Application US/09764855  
; Patent No. US20020119919A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P110  
; CURRENT APPLICATION NUMBER: US/09/764,855  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 334  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 99  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-855-99

Query Match 6.2%; Score 6; DB 10; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19  
Db 45 SETLSQ 50

RESULT 34

US-10-091-504-1022  
; Sequence 1022, Application US/10091504  
; Publication No. US20030059908A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC007C1  
; CURRENT APPLICATION NUMBER: US/10/091,504  
; CURRENT FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 2442  
; Prior Application removed - See File Wrapper or Palm  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1022  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-091-504-1022

Query Match  
Best Local Similarity 6.2%; Score 6; DB 9; Length 97;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TELRKK 25  
Db 74 TELRKK 79

RESULT 35  
US-09-764-869-1022  
; Sequence 1022, Application US/09764869  
; Patent No. US20020061521A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC007  
; CURRENT APPLICATION NUMBER: US/09/764,869  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2442  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1022  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-869-1022

Query Match  
Best Local Similarity 6.2%; Score 6; DB 10; Length 97;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TELRKK 25  
Db 74 TELRKK 79

RESULT 36  
US-09-925-299-774  
; Sequence 774, Application US/09925299  
; Publication No. US20030040617A9  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 774  
; LENGTH: 101

; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (69)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (98)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-774

Query Match  
Best Local Similarity 6.2%; Score 6; DB 9; Length 101;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 PPSPE 71  
Db 52 PPSPE 57

RESULT 37  
US-09-925-299-774  
; Sequence 774, Application US/09925299  
; Patent No. US20020055627A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 774  
; LENGTH: 101  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (69)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (98)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-774

Query Match  
Best Local Similarity 6.2%; Score 6; DB 10; Length 101;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 PPSPE 71  
Db 52 PPSPE 57

RESULT 38  
US-10-043-487-333  
; Sequence 333, Application US/10043487  
; Publication No. US20030055220A1  
; GENERAL INFORMATION:  
; APPLICANT: HYBRIGENICS  
; APPLICANT: Pierre, LEGRAIN  
; TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypep  
; FILE REFERENCE: B4778A  
; CURRENT APPLICATION NUMBER: US/10/043,487  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/261,130  
; PRIOR FILING DATE: 2001-01-12  
; NUMBER OF SEQ ID NOS: 561



; NUMBER OF SEQ ID NOS: 104  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 29  
; LENGTH: 141  
; TYPE: PRT  
; ORGANISM: Canis familiaris  
US-09-828-995B-29

Query Match  
Best Local Similarity 6.2%; Score 6; DB 9; Length 141;  
Matches 6; Conservativity 100.0%; Pred. No. 1.8e+02;  
Mismatches 0; Indels 0; Gaps 0;

QY 66 PPSPKE 71  
Db 43 PPSPKE 48

RESULT 42  
US-09-864-761-36181  
; Sequence 36181, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 36181  
; LENGTH: 141  
; TYPE: PRT  
; ORGANISM: Homo sapiens

; FEATURE:  
; OTHER INFORMATION: MAP TO AL023333.1  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.4  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.3  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.2  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9  
US-09-864-761-36181

Query Match  
Best Local Similarity 6.2%; Score 6; DB 10; Length 141;  
Matches 6; Conservativity 100.0%; Pred. No. 1.8e+02;  
Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERK 28  
Db 44 RKKERK 49

RESULT 43  
US-09-789-561-88  
; Sequence 88, Application US/09789561  
; Patent No. US20020064818A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 52 Human secreted proteins  
; FILE REFERENCE: PZ043P1  
; CURRENT APPLICATION NUMBER: US/09/789,561  
; PRIOR FILING DATE: 2001-02-22  
; PRIOR APPLICATION NUMBER: PCT/US00/24008  
; PRIOR FILING DATE: 2000-08-31  
; PRIOR APPLICATION NUMBER: 60/152,317  
; PRIOR FILING DATE: 1999-09-03  
; PRIOR APPLICATION NUMBER: 60/152,315  
; PRIOR FILING DATE: 1999-09-03  
; NUMBER OF SEQ ID NOS: 194  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 88  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-789-561-88

Query Match  
Best Local Similarity 6.2%; Score 6; DB 10; Length 151;  
Matches 6; Conservativity 100.0%; Pred. No. 1.9e+02;  
Mismatches 0; Indels 0; Gaps 0;

QY 63 LLCPPS 68  
Db 143 LLCPPS 148

RESULT 44  
US-09-321-801-17  
; Sequence 17, Application US/09321801  
; Patent No. US20020115176A1  
; GENERAL INFORMATION:  
; APPLICANT: Lanfear, Jeremy  
; APPLICANT: Robas, Nicola M.  
; TITLE OF INVENTION: PHOSPHODIESTERASE ENZYMES  
; FILE REFERENCE: PC9477A  
; CURRENT APPLICATION NUMBER: US/09/321,801  
; PRIOR FILING DATE: 1999-05-27  
; EARLIER APPLICATION NUMBER: 9826777.6  
; EARLIER FILING DATE: 1998-12-04  
; EARLIER APPLICATION NUMBER: 9823882.7  
; EARLIER FILING DATE: 1998-10-30  
; EARLIER APPLICATION NUMBER: 9811500.9  
; EARLIER FILING DATE: 1998-05-28



us-09-854-133-586.olig.rapb

Tue May 13 12:12:43 2003

EARLIER APPLICATION NUMBER: 9908247.1  
 EARLIER FILING DATE: 1998-04-09  
 EARLIER APPLICATION NUMBER: 9910801.1  
 EARLIER FILING DATE: 1999-05-10  
 NUMBER OF SEQ ID NOS: 19  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 17  
 LENGTH: 160  
 TYPE: PRT  
 ORGANISM: Rat  
 US-09-321-801-17

Query Match 6.2%; Score 6; DB 10; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 PSPKEV 72  
 Db 68 PSPKEV 73

RESULT 45

US-10-211-962-77  
 ; Sequence 77, Application US/10211962  
 ; Publication NO. US20030082640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Herz, Joachim  
 ; APPLICANT: Gotthardt, Michael  
 ; TITLE OF INVENTION: LDL Receptor Signaling Pathways  
 ; FILE REFERENCE: UTSW0708  
 ; CURRENT APPLICATION NUMBER: US/10/211,962  
 ; CURRENT FILING DATE: 2002-08-01  
 ; PRIOR APPLICATION NUMBER: US/09/562,737  
 ; PRIOR FILING DATE: 2000-05-01  
 ; NUMBER OF SEQ ID NOS: 132  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 77  
 ; LENGTH: 163  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 ; OTHER INFORMATION: Sequence  
 US-10-211-962-77

Query Match 6.2%; Score 6; DB 9; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ASLGDS 14  
 Db 13 ASLGDS 18

Search completed: May 11, 2003, 20:21:13  
 Job time : 34.7611 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:07:52 ; Search time 30.0442 Seconds  
(without alignments)  
310.377 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 97  
Sequence: 1 EVEVSRDHASLGDSSETLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	8.2	67	2 A70375	ribosomal protein
2	7	7.2	129	2 B71554	ribosomal protein
3	7	7.2	133	2 H81705	ribosomal protein
4	7	7.2	134	2 B72101	ribosomal protein
5	7	7.2	134	2 F86521	ribosomal protein
6	7	7.2	197	2 T17106	hypothetical prote
7	7	7.2	310	1 E64923	probable transcrip
8	7	7.2	310	2 H90924	probable transcrip
9	7	7.2	310	2 D85773	probable transcrip
10	7	7.2	310	2 AI0695	hypothetical prote
11	7	7.2	318	2 C84651	hypothetical prote
12	7	7.2	340	2 C75004	hypothetical prote
13	7	7.2	356	2 S55181	hypothetical prote
14	7	7.2	383	2 F82871	conserved hypothet
15	7	7.2	394	2 C84905	probable extensin
16	7	7.2	409	2 AF3271	probable allantoin
17	7	7.2	463	2 AH2025	hypothetical prote
18	7	7.2	483	2 S41689	cytochrome-c oxida
19	7	7.2	483	2 F71619	hypothetical prote
20	7	7.2	557	2 T04465	hypothetical prote
21	7	7.2	560	2 F70719	hypothetical prote
22	7	7.2	583	2 S46265	cysteine proteinas
23	7	7.2	628	2 T01467	hypothetical prote
24	7	7.2	648	2 C69423	DNA helicase homol
25	7	7.2	651	2 C86333	hypothetical prote
26	7	7.2	703	2 G75638	nodulation protein
27	7	7.2	777	2 T04768	hypothetical prote
28	7	7.2	858	2 A46613	protein 4.1, p4.1
29	7	7.2	1257	2 T28937	hypothetical prote

30	7	7.2	1286	2 T16507	hypothetical prote
31	7	7.2	1435	2 S69632	regulatory protein
32	7	7.2	1551	2 F86342	F9H16.4 protein -
33	7	7.2	3844	2 T18402	asparagine/asparta
34	6	6.2	63	2 D69398	hypothetical prote
35	6	6.2	74	2 AG3625	hypothetical prote
36	6	6.2	80	2 G84773	hypothetical prote
37	6	6.2	89	2 B84145	hypothetical prote
38	6	6.2	100	2 T04102	hypothetical prote
39	6	6.2	103	2 T17566	smt3 protein - ric
40	6	6.2	110	2 T07618	hypothetical prote
41	6	6.2	116	2 B40617	cold stress protei
42	6	6.2	116	2 H72386	transcription repr
43	6	6.2	121	2 D72336	mazG-related prote
44	6	6.2	121	2 F84868	transcription regu
45	6	6.2	124	2 T37059	hypothetical prote
46	6	6.2	129	2 F83052	hypothetical prote
47	6	6.2	132	2 T04105	ribosome-binding f
48	6	6.2	132	2 B69454	probable signal re
49	6	6.2	136	2 B82298	conserved hypothet
50	6	6.2	137	2 S73212	ribosome-binding f
51	6	6.2	142	2 T46042	ribosomal protein
52	6	6.2	143	2 T46122	bZIP transcription
53	6	6.2	144	2 T18867	30S ribosomal prot
54	6	6.2	156	2 G69233	hypothetical prote
55	6	6.2	162	2 H70314	N-terminal acetyl
56	6	6.2	166	2 D37844	hypothetical prote
57	6	6.2	167	2 T10639	baIE 19.5K protein
58	6	6.2	169	2 G75319	hypothetical prote
59	6	6.2	173	2 T45993	hypothetical prote
60	6	6.2	184	2 I38910	hypothetical 19.8K
61	6	6.2	189	1 LNRBC1	pulmonary surfacta
62	6	6.2	201	2 G70156	hypothetical prote
63	6	6.2	205	2 E64621	conserved hypothet
64	6	6.2	205	2 T00865	hypothetical prote
65	6	6.2	206	2 C70950	hypothetical prote

## ALIGNMENTS

## RESULT 1

A70375 ribosomal protein S21 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 13-Aug-1999

C:Accession: A70375

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.F. V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98195666; PMID:9537320

A:Accession: A70375

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-67 <AQF>

A:Cross-references: GB:AE000711; NID:g2983401; PIDN:AAC06990.1; PID:g2983406; GB:AE000711

A:Experimental source: strain VF5

C:Genetics:

A:Gene: rpsu

C:Superfamily: Escherichia coli ribosomal protein S21

Query Match 8.2%; Score 8; DB 2; Length 67;  
Best Local Similarity 100.0%; Pred. No. 0.33;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ERKKKRER 33

Db 43 ERKKKRER 50

## RESULT 2

B71554

```

ribosomal protein S9 [similarity] - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 03-May-2002
C;Accession: B71554
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
A;Reference number: A71570; MUID:99000809; PMID:9784136
A;Accession: B71554
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-129 <ARN>
A;Cross-references: GB:AE001286; GB:AE001273; NID:g3328516; PIDN:AAC67717.1; PID:g332852
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: rs9
C;Superfamily: Escherichia coli ribosomal protein S9

Query Match          7.2%  Score 7;  DB 2;  Length 129;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29
Db 108 RKKERKK 114

RESULT 3
H81705
ribosomal protein S9 TC0402 [imported] - Chlamydia muridarum (strain Nigg)
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C;Accession: H81705
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: H81705
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-133 <TET>
A;Cross-references: GB:AE002307; GB:AE002160; NID:g7190442; PIDN:AAF39259.1; PID:g719044
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0402
C;Superfamily: Escherichia coli ribosomal protein S9

Query Match          7.2%  Score 7;  DB 2;  Length 133;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29
Db 112 RKKERKK 118

RESULT 4
B72101
ribosomal protein S9 CP0516 [similarity] - Chlamydia pneumoniae (strains CWL029 and
C;Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 03-May-2002
C;Accession: B72101; F81567
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A;Reference number: A72000; MUID:99206606; PMID:10192388
A;Accession: B72101
A;Molecule type: DNA
A;Residues: 1-134 <ARN>
A;Cross-references: GB:AE001610; GB:AE001363; NID:g4376515; PIDN:AAD18399.1; PID:g437652
A;Experimental source: strain CWL029
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,

```

```

, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Sal
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae A
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: F81567
A;Molecule type: DNA
A;Residues: 1-134 <REA>
A;Cross-references: GB:AE002212; GB:AE002161; NID:g7189430; PIDN:AAF38343.1; PID:g
A;Experimental source: strain AR39, HL cells
C;Genetics:
A;Gene: rs9; CP0516
C;Superfamily: Escherichia coli ribosomal protein S9

Query Match          7.2%  Score 7;  DB 2;  Length 134;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29
Db 113 RKKERKK 119

RESULT 5
F86521
ribosomal protein S9 [similarity] - Chlamydia pneumoniae (strain J138)
C;Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 03-May-2002
C;Accession: F86521
R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T
Nucleic Acids Res. 28, 2311-2314, 2000
A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A;Reference number: A86491; MUID:20330349; PMID:10871362
A;Accession: F86521
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-134 <STO>
A;Cross-references: GB:BA000008; NID:g8978619; PIDN:BAA98456.1; GSPDB:GN00142
A;Experimental source: strain J138
C;Genetics:
A;Gene: rs9
C;Superfamily: Escherichia coli ribosomal protein S9

Query Match          7.2%  Score 7;  DB 2;  Length 134;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29
Db 113 RKKERKK 119

RESULT 6
T17106
hypothetical protein PAFD103 - apple tree (fragment)
C;Species: Malus domestica (apple tree)
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C;Accession: T17106
R;Dong, Y.H.; Janssen, B.J.; Bielecki, L.L.; Atkinson, R.G.; Morris, B.A.; Gardner,
J. Am. Soc. Hort. Sci. 122, 752-757, 1997
A;Title: Isolating and characterizing genes differentially expressed early in apple
A;Reference number: Z18681
A;Accession: T17106
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-197 <DON>
A;Cross-references: EMBL:U80270; NID:g1732362; PID:g1732363
A;Experimental source: strain Granny Smith
C;Genetics:
A;Note: PAFD103

Query Match          7.2%  Score 7;  DB 2;  Length 197;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



Tue May 13 12:12:44 2003

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C;Accession: D85773  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Ma  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apod  
Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: D85773

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-310 &lt;STO&gt;

A;Cross-references: GB:AE005174; NID:gl2515657; PIDN:AAG56648.1; GSPDB:GN00145; UWG

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: ydhB

C;Superfamily: probable transcription regulator ybbs

Query Match 7.2%; Score 7; DB 2; Length 310;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSETL 17

|||||

Db 288 LGDSETL 294

## RESULT 10

AI0695

probable transcription regulator STY1693 [imported] - Salmonella enterica subsp. en

C;Species: Salmonella enterica subsp. enterica serovar Typhi

A;Note: this species has also been called Salmonella typhi

C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001

C;Accession: AI0695

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Chur  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Fa  
S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens

A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica

A;Reference number: AB0502; PMID:11677608

A;Accession: AI0695

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-310 &lt;PAR&gt;

A;Cross-references: GB:AL513382; PIDN:CAD01938.1; PID:gl6502780; GSPDB:GN00176

C;Genetics:

A;Gene: STY1693

C;Superfamily: probable transcription regulator ybbs

Query Match 7.2%; Score 7; DB 2; Length 310;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSETL 17

|||||

Db 288 LGDSETL 294

## RESULT 11

C84651

hypothetical protein At2g25670 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C;Accession: C84651

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii,  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tall  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ven  
Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: C84651

A;Status: preliminary

A;Molecule type: DNA

QY 24 KKERKK 30

|||||

Db 54 KKERKK 60

## RESULT 7

E64923

probable transcription regulator ydhB - Escherichia coli (strain K-12)

C;Species: Escherichia coli

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 01-Mar-2002

C;Accession: E64923

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: E64923

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-310 &lt;BLAT&gt;

A;Cross-references: GB:AE000261; GB:U00096; NID:gl787945; PIDN:AAC74731.1; PID:gl787949;

A;Experimental source: strain K-12, substrain MGL655

C;Genetics:

A;Gene: ydhB

C;Superfamily: probable transcription regulator ybbs

C;Keywords: DNA binding; transcription regulation

F;18-48/Region: regulatory protein lysR motif

F;19-38/Region: helix-turn-helix motif

Query Match 7.2%; Score 7; DB 1; Length 310;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSETL 17

|||||

Db 288 LGDSETL 294

## RESULT 8

H90924

probable transcription regulator LYSR-type [imported] - Escherichia coli (strain O157:H7)

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C;Accession: H90924

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A;Reference number: A9629; MUID:21156231; PMID:11258796

A;Accession: H90924

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-310 &lt;HAY&gt;

A;Cross-references: GB:BA000007; PIDN:BA035791.1; PID:gl13361835; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain RIMD 0509952

C;Genetics:

A;Gene: ECs2368

C;Superfamily: probable transcription regulator ybbs

Query Match 7.2%; Score 7; DB 2; Length 310;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSETL 17

|||||

Db 288 LGDSETL 294

## RESULT 9

D85773

probable transcription regulator LYSR-type ydhB [imported] - Escherichia coli (strain O1

C;Species: Escherichia coli

A;Residues: 1-318 <STO>  
A;Cross-references: GB:AE002093; NID:g4874305; PIDN:AAD31367.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g25670  
A;Map position: 2

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 318;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 24 KKKKKK 30  
Db 160 KKKKKK 166

## RESULT 12

C75004

hypothetical protein PAB1036 - Pyrococcus abyssi (strain Orsay)  
C;Species: Pyrococcus abyssi

C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C;Accession: C75004

R;anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999

A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure

A;Reference number: A75001

A;Accession: C75004

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-340 <KAW>

A;Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50473.1; PID:g545898

A;Experimental source: strain Orsay

C;Genetics:

A;Gene: PAB1036

C;Superfamily: Pyrococcus abyssi hypothetical protein PAB1036

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 340;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 KKKKKK 31

Db 287 KKKKKK 293

## RESULT 13

S55181

hypothetical protein YJL131c - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein J0682

C;Species: Saccharomyces cerevisiae

C;Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 19-Apr-2002

C;Accession: S55181; S05768; S56913; S71666; PQ0003

R;Katsoulou, C.; Tzermia, M.; Alexandraki, D.

submitted to the EMBL Data Library, May 1995

A;Description: The complete sequence of a 40.7 kb segment located on the left arm of yeast hypothetical proteins.

A;Reference number: S55159

A;Accession: S55181

A;Molecule type: DNA

A;Residues: 1-356 <KAT>

A;Cross-references: EMBL:X87371; NID:g854542; PIDN:CAA60824.1; PID:g854565

R;Souciet, J.L.; Nagy, M.; le Gouar, M.; Lacroute, F.; Potier, S.

Gene 79, 59-70, 1989

A;Title: Organization of the yeast URA2 gene: identification of a defective dihydroorotase

A;Reference number: S05766; MUID:89378778; PMID:2570735

A;Accession: S05768

A;Molecule type: DNA

A;Residues: 1-124 <SOU>

A;Cross-references: EMBL:M27174

R;Katsoulou, C.; Tzermia, M.; Alexandraki, D.

submitted to the Protein Sequence Database, September 1995

A;Reference number: S56912

A;Accession: S56913

A;Molecule type: DNA

A;Residues: 1-356 <KAW>

A;Cross-references: EMBL:Z49406; NID:gl008333; PIDN:CAA89426.1; PID:gl008334; MIPS

R;Katsoulou, C.; Tzermia, M.; Tavernarakis, N.; Alexandraki, D.

Yeast 12, 787-797, 1996

A;Title: Sequence analysis of a 40.7 kb segment from the left arm of yeast chromosome XI.

A;Reference number: S71643; MUID:96408771; PMID:8813765

A;Accession: S71666

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-356 <KAF>

A;Cross-references: EMBL:X87371; NID:g854542; PIDN:CAA60824.1; PID:g854565

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995

C;Genetics:

A;Cross-references: SGD:S0003667

A;Map position: 10L

## Query Match

Best Local Similarity 7.2%; Score 7; DB 2; Length 356;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 ELKKKK 27

Db 334 ELKKKK 340

## RESULT 14

F82871

conserved hypothetical membrane lipoprotein UU602 [imported] - Ureaplasma urealyticum

C;Species: Ureaplasma urealyticum

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C;Accession: F82871

R;Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.

submitted to GenBank, February 2000

A;Description: The complete sequence of Ureaplasma urealyticum: Alternate views of

A;Reference number: A82870

A;Accession: F82871

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-383 <GLA>

A;Cross-references: GB:AE002158; GB:AF222894; NID:g6899599; PIDN:AAF31016.1; GSPDB:

A;Experimental source: serovar 3; biovar 1

C;Genetics:

A;Gene: UU602

A;Genetic code: SGC3

## Query Match

Best Local Similarity 7.2%; Score 7; DB 2; Length 383;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 KKKKKK 30

Db 30 KKKKKK 36

## RESULT 15

C84905

probable extensin [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C;Accession: C84905

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Vent

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: C84905

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-394 <STO>

A;Cross-references: GB:AE002093; NID:g3831447; PIDN:AAC69930.1; GSPDB:GN00139

C;Genetics:

Tue May 13 12:12:44 2003

A;Gene: At2g46630  
A;Map position: 2

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 394;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 PPSPKEV 72  
| | | | |  
Db 124 PPSPKEV 130

RESULT 16  
AF3271  
probable allantoin permease [imported] - Brucella melitensis (strain 16M)  
C;Species: Brucella melitensis  
C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C;Accession: AF3271  
R;DelVecchio, V.G.; Kapatral, R.J.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, S.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, M.; Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A;Reference number: AD3252; PMID:11756688  
A;Accession: AF3271  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-409 <KUR>  
A;Cross-references: GB:AE008917; PIDN:AAL51337.1; PID:gl7982035; GSPDB:GN00190  
A;Experimental source: strain 16M  
C;Genetics:  
A;Gene: BMEI0155  
A;Map position: 1

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 409;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 SQTELK 24  
| | | | |  
Db 396 SQTELK 402

RESULT 17  
AH2025  
hypothetical protein all1758 [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp.  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C;Accession: AH2025  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.; DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AH2025  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-463 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BA073457.1; PID:gl7130848; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: all1758

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 463;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 ELKKER 27  
| | | | |  
Db 199 ELKKER 205

RESULT 18

S41689  
cytochrome-c oxidase (EC 1.9.3.1) chain I - Theileria parva mitochondrion (fragment)  
C;Species: mitochondrion Theileria parva  
C;Date: 31-Dec-1993 #sequence\_revision 02-Aug-1994 #text\_change 07-Dec-1999  
C;Accession: S41689; S40154  
R;Kairo, A.; Fairlamb, A.H.; Gobright, E.; Nene, V.  
EMBO J. 13, 898-905, 1994  
A;Title: A 7.1 kb linear DNA molecule of Theileria parva has scrambled rDNA sequence  
A;Reference number: S41689; MUID:94155854; PMID:8112303  
A;Accession: S41689  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-483 <KAI>  
A;Cross-references: EMBL:Z23263; NID:g437862; PIDN:CAA80798.1; PID:g437863  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1993  
C;Genetics:  
A;Genome: mitochondrion  
A;Genetic code: SGC6  
C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
C;Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane  
transmembrane protein  
F;18-466/Domains: cytochrome-c oxidase chain I homology <COI>  
F;71,386/Binding site: heme a iron (His) (axial ligands) #status predicted  
F;248-252/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
F;252/Binding site: oxygen (Tyr) #status predicted  
F;384/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 7.2%; Score 7; DB 2; Length 483;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 GIDFIIF 46  
| | | | |  
Db 150 GIDFIIF 156

RESULT 19  
F71619  
hypothetical protein PFB0235w - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C;Accession: F71619  
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, T.; Science 282, 1126-1132, 1998  
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A;Reference number: A71600; MUID:99021743; PMID:9804551  
A;Accession: F71619  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-483 <GAR>  
A;Cross-references: GB:AE001382; GB:AE001362; NID:g3845130; PIDN:AAC71836.1; PID:g3845130  
A;Experimental source: clone 3D7  
C;Genetics:  
A;Gene: PFB0235w

Query Match 7.2%; Score 7; DB 2; Length 483;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 30  
| | | | |  
Db 459 KKERKK 465

RESULT 20  
T04465  
hypothetical protein F4D11.190 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cross)  
C;Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999  
C;Accession: T04465  
R;Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Anson, W.; Hoheisel, J.; Mewes, F.; submitted to the Protein Sequence Database, April 1998

A;Reference number: Z15360  
A;Accession: T04465  
A;Molecule type: DNA  
A;Residues: 1-557 <BEV>  
A;Cross-references: EMBL:AL022537  
A;Experimental source: cultivar Columbia; BAC clone F4D11  
C;Genetics:  
A;Map position: 4  
A;Introns: 107/3; 196/1; 315/2; 388/3; 497/2  
A;Note: F4D11.190

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 557;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKKK 30  
Db 163 KKERKKK 169

#### RESULT 21

F70719  
hypothetical protein Rv0976c - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C;Accession: F70719  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: F70719  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-560 <COL>  
A;Cross-references: GB:Z79700; GB:AL123456; NID:g3261628; PIDN:CAB02013.1; PID:el300054;  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: Rv0976c

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 560;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 REMLTGG 81  
Db 18 REMLTGG 24

#### RESULT 22

S46265  
cysteine proteinase - Plasmodium vivax  
C;Species: Plasmodium vivax  
C;Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 29-Sep-1999  
C;Accession: S46265  
R;Rosenthal, P.J.; Ring, C.S.; Chen, X.; Cohen, F.E.  
J. Mol. Biol. 241, 312-316, 1994  
A;Title: Characterization of a Plasmodium vivax cysteine proteinase gene identifies unique  
A;Reference number: S46265; MUID:94334995; PMID:8057374  
A;Accession: S46265  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-583 <ROS>  
A;Cross-references: EMBL:L26362; NID:g431433; PIDN:AAA60368.1; PID:g431434  
C;Superfamily: trophozoite cysteine proteinase

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 583;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ERKKKRE 32

Db 175 ERKKKRE 181

#### RESULT 23

T01467  
hypothetical protein T24H24.5 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 04-Mar-2000  
C;Accession: T01467  
R;Courtney, L.; Stoneking, T.; Langston, Y.; Mead, K.  
submitted to the EMBL Data Library, August 1998  
A;Description: The sequence of A. thaliana T24H24.  
A;Reference number: Z14333  
A;Accession: T01467  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-628 <COU>  
A;Cross-references: EMBL:AF075598; NID:g3293581; PID:g3377830  
A;Experimental source: cultivar Columbia  
C;Genetics:  
A;Map position: 4  
A;Introns: 55/2; 216/1; 289/1; 363/3; 413/3  
A;Note: T24H24.5  
C;Superfamily: Arabidopsis thaliana hypothetical protein T24H24.5

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 628;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TLSQTEL 22  
Db 222 TLSQTEL 228

#### RESULT 24

C69423  
DNA helicase homolog - Archaeoglobus fulgidus  
C;Species: Archaeoglobus fulgidus  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999  
C;Accession: C69423  
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Do  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness,  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arc  
A;Reference number: A69250; MUID:98049343; PMID:9389475  
A;Accession: C69423  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-648 <KLE>  
A;Cross-references: GB:AE001009; GB:AE000782; NID:g2689332; PIDN:AAB89860.1; PID:g264  
C;Superfamily: probable DNA helicase MJ0104

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 648;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 GDSETLS 18  
Db 609 GDSETLS 615

#### RESULT 25

C86333  
hypothetical protein T20H2.23 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C;Accession: C86333  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar,



```

anssen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Marzialli,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C86333
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-651 <STO>
A;Cross-references: GB:AE005172; NID:g8778999; PIDN:AAF79914.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match      7.2%  Score 7;  DB 2;  Length 651;
Best Local Similarity 100.0%; Pred. No. 26;
Matches      7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      28 KKKRERK 34
      |||||
Db      507 KKKRERK 513

RESULT 26
G75638
modulation protein-related protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: G75638
R;White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: G75638
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-703 <WHI>
A;Cross-references: GB:AE001827; NID:g6460959; PIDN:AAF12674.1; PID:g6460971; TIGR:DRC00
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRC0037
A;Map position: plasmid
A;Genome: plasmid
A;Note: plasmid CPl

Query Match      7.2%  Score 7;  DB 2;  Length 703;
Best Local Similarity 100.0%; Pred. No. 28;
Matches      7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      12 GDSETLS 18
      |||||
Db      165 GDSETLS 171

RESULT 27
T04768
hypothetical protein T16H5.230 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T04768
R;Bevan, M.; De Haan, M.; Maarse, A.C.; Grivell, L.A.; Bancroft, I.; Mewes, H.W.; Mayer,
submitted to the Protein Sequence Database, June 1998
A;Reference number: Z15383
A;Accession: T04768
A;Molecule type: DNA
A;Residues: 1-777 <BEV>
A;Cross-references: EMBL:AL024486
A;Experimental source: cultivar Columbia; BAC clone T16H5

anssen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Marzialli,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C86333
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-651 <STO>
A;Cross-references: GB:AE005172; NID:g8778999; PIDN:AAF79914.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match      7.2%  Score 7;  DB 2;  Length 651;
Best Local Similarity 100.0%; Pred. No. 26;
Matches      7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      28 KKKRERK 34
      |||||
Db      507 KKKRERK 513

RESULT 26
G75638
modulation protein-related protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: G75638
R;White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: G75638
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-703 <WHI>
A;Cross-references: GB:AE001827; NID:g6460959; PIDN:AAF12674.1; PID:g6460971; TIGR:DRC00
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRC0037
A;Map position: plasmid
A;Genome: plasmid
A;Note: plasmid CPl

Query Match      7.2%  Score 7;  DB 2;  Length 703;
Best Local Similarity 100.0%; Pred. No. 28;
Matches      7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      12 GDSETLS 18
      |||||
Db      165 GDSETLS 171

RESULT 27
T04768
hypothetical protein T16H5.230 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T04768
R;Bevan, M.; De Haan, M.; Maarse, A.C.; Grivell, L.A.; Bancroft, I.; Mewes, H.W.; Mayer,
submitted to the Protein Sequence Database, June 1998
A;Reference number: Z15383
A;Accession: T04768
A;Molecule type: DNA
A;Residues: 1-777 <BEV>
A;Cross-references: EMBL:AL024486
A;Experimental source: cultivar Columbia; BAC clone T16H5

C;Genetics:
A;Map position: 4
A;Introns: 371/1
A;Note: T16H5.230

Query Match      7.2%  Score 7;  DB 2;  Length 777;
Best Local Similarity 100.0%; Pred. No. 30;
Matches      7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      86 ATRSHLG 92
      |||||
Db      455 ATRSHLG 461

RESULT 28
A46613
protein 4.1, p4.1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 25-Aug-1995
C;Accession: A46613
R;Huang, J.P.; Tang, C.J.; Kou, G.H.; Marchesi, V.T.; Benz Jr., E.J.; Tang, T.K.
J. Biol. Chem. 268, 3758-3766, 1993
A;Title: Genomic structure of the locus encoding protein 4.1. Structural basis for c
A;Reference number: A46613; MUID:93155238; PMID:8429050
A;Accession: A46613
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-858 <HUA>
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIN:124466, NCBIIP:124467)
C;Superfamily: membrane protein 4.1; protein 4.1 membrane-binding domain homology
F;213-488/Domain: protein 4.1 membrane-binding domain homology <B41>

Query Match      7.2%  Score 7;  DB 2;  Length 858;
Best Local Similarity 100.0%; Pred. No. 33;
Matches      7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      27 RKKRRER 33
      |||||
Db      641 RKKRRER 647

RESULT 29
T28937
hypothetical protein C52B9.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T28937
R;Nelson, J.
submitted to the EMBL Data Library, July 1996
A;Description: The sequence of C. elegans cosmid C52B9.
A;Reference number: Z20545
A;Accession: T28937
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1257 <NEL>
A;Cross-references: EMBL:U64598; PIDN:AAC47974.1; GSPDB:GN00028; CESP:C52B9.8
A;Experimental source: strain Bristol N2; clone C52B9
C;Genetics:
A;Gene: CESP:C52B9.8
A;Map position: X
A;Introns: 15/2; 321/3; 450/3; 596/2; 776/2; 823/2; 871/3

Query Match      7.2%  Score 7;  DB 2;  Length 1257;
Best Local Similarity 100.0%; Pred. No. 45;
Matches      7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      19 QTELKK 25
      |||||
Db      55 QTELKK 61

RESULT 30

```

Tl6507  
hypothetical protein F59A6.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: Tl6507  
R:Nhan, M.  
submitted to the EMBL Data Library, December 1995  
A:Description: The sequence of C. elegans cosmid F59A6.  
A:Reference number: Z18526  
A:Accession: Tl6507  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1286 <NHA>  
A:Cross-references: EMBL:U41994; NID:g1123047; PID:g1123050; PIDN:AAA83454.1; CESP:F59A6  
A:Gene: CESP:F59A6.5  
A:Introns: 35/3; 335/3; 685/3; 973/3; 1097/3

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 1286;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 ASLGDSE 15  
|||||  
Db 317 ASLGDSE 323

RESULT 31  
S69632  
regulatory protein SPP41 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YDR464w  
C:Species: Saccharomyces cerevisiae  
C:Date: 22-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 23-Mar-2001  
R:Dieitrich, F.S.  
submitted to the EMBL Data Library, August 1995  
A:Description: The sequence of S. cerevisiae cosmids 9410, 8035, 8166, and 9787.  
A:Reference number: S69554  
A:Accession: S69632  
A:Molecule type: DNA  
A:Residues: 1-1435 <DIE>  
A:Cross-references: EMBL:U33050; NID:g927726; PID:g927729; MIPS:YDR464w  
R:Maddock, J.R.; Weidenhammer, E.M.; Adams, C.C.; Lunz, R.L.; Woolford Jr., J.L.  
Genetics 136, 833-847, 1994  
A:Title: Extragenic suppressors of Saccharomyces cerevisiae prp4 mutations identify a ne  
A:Reference number: S47864; MUID:94274035; PMID:8005438  
A:Accession: S47864  
A:Molecule type: DNA  
A:Residues: 1-1394, 'R' <MAD>  
A:Cross-references: EMBL:U03673; NID:g435022; PIDN:AAA20494.1; PID:g435023  
C:Genetics:  
A:Gene: SGD:SPP41  
A:Cross-references: SGD:S0002872; MIPS:YDR464w  
A:Map position: 4R  
C:Keywords: nucleus

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 1435;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 24 KKERKKK 30  
|||||  
Db 695 KKERKKK 701  
RESULT 32  
F86342  
F9H16.4 protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Nov-2001  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Ma  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: F86342  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1551 <STO>  
A:Cross-references: GB:AE005172; NID:g4836891; PIDN:AAD30594.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 1551;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 26 ERKKKRE 32  
|||||  
Db 1304 ERKKKRE 1310

RESULT 33  
T18402  
asparagine/aspartate rich protein - malaria parasite (Plasmodium falciparum) (fragm  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
R:Barale, J.C.; Candelle, D.; Attal-Bonnefoy, G.; Dehoux, P.; Bonnefoy, S.; Ridley,  
Infect. Immun. 65, 3003-3010, 1997  
A:Title: Plasmodium falciparum AARP1, a giant protein containing repeated motifs r  
A:Reference number: Z18929; MUID:97378065; PMID:9234746  
A:Accession: T18402  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-3844 <BAR>  
A:Cross-references: EMBL:Y08926; NID:el154302; PID:e311435; PIDN:CAA70130.1  
C:Genetics:  
A:Gene: aarp1

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 3844;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 24 KKERKKK 30  
|||||  
Db 3134 KKERKKK 3140

RESULT 34  
D69398  
hypothetical protein AF1189 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C:Accession: D69398  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch  
A:Reference number: A69250; MUID:98049343; PMID:9389475  
A:Accession: D69398  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-63 <KLE>  
A:Cross-references: GB:AE001022; GB:AE000782; NID:g2689345; PIDN:AAB90067.1; PID:g264

C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C;Accession: B84145  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans  
A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: B84145  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-89 <STO>  
A;Cross-references: GB:AP001520; GB:BA000004; NID:g10176401; PIDN:BAB07681.1; GSPDB:  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH3962

Query Match 6.2%; Score 6; DB 2; Length 89;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 KKKRER 33  
|||||  
Db 84 KKKRER 89

RESULT 38

T04102  
smt3 protein - rice  
C;Species: Oryza sativa (rice)  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-Jul-2000  
C;Accession: T04102  
R;Lapenta, V.; Chiurazzi, P.; Vanderspek, P.; Pizzuti, A.; Hanaoka, F.; Brahe, C.  
Genomics 40, 362-366, 1997  
A;Title: Smt3a, a human homolog of the s-cerevisiae-smt3 gene, maps to chromosome-2  
A;Reference number: Z05239; MUID:97237059; PMID:9119407  
A;Accession: T04102  
A;Status: translated from GB/EMBL/DBBJ  
A;Molecule type: mRNA  
A;Residues: 1-100 <LAP>  
A;Cross-references: EMBL:X99608; NID:g1668772; PIDN:CAA67922.1; PID:g1668773  
C;Genetics:  
A;Gene: smt3  
C;Superfamily: yeast SMT3 protein

Query Match 6.2%; Score 6; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 TGGCLP 84  
|||||  
Db 94 TGGCLP 99

RESULT 39

T17566  
hypothetical protein a76L - Chlorella virus PBCV-1  
C;Species: Chlorella virus PBCV-1  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C;Accession: T17566  
R;Graves, M.V.; Van Etten, J.L.  
submitted to the EMBL Data Library, May 1999  
A;Reference number: Z18806  
A;Accession: T17566  
A;Status: preliminary; translated from GB/EMBL/DBBJ  
A;Molecule type: DNA  
A;Residues: 1-103 <GRA>  
A;Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96444.1  
A;Experimental source: specific host Chlorella strain NC64A  
C;Genetics:  
A;Gene: a76L  
C;Superfamily: Chlorella virus PBCV-1 hypothetical protein a76L

Query Match 6.2%; Score 6; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 66;

Query Match 6.2%; Score 6; DB 2; Length 63;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ERKKR 31  
|||||  
Db 51 ERKKR 56

RESULT 35  
AG3625  
hypothetical protein BMEII0928 [imported] - Brucella melitensis (strain 16M)  
C;Species: Brucella melitensis  
C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C;Accession: AG3625  
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A;Reference number: AD3252; PMID:11756688  
A;Accession: AG3625  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-74 <KUR>  
A;Cross-references: GB:AE008918; PIDN:AAL54170.1; PID:g17985136; GSPDB:GN00191  
A;Experimental source: strain 16M  
C;Genetics:  
A;Gene: BMEII0928  
A;Map position: II

Query Match 6.2%; Score 6; DB 2; Length 74;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 LLFSHH 57  
|||||  
Db 63 LLFSHH 68

RESULT 36

G84773  
hypothetical protein At2g35850 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: G84773  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: G84773  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-80 <STO>  
A;Cross-references: GB:AE002093; NID:g4510384; PIDN:AAD21472.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g35850  
A;Map position: 2

Query Match 6.2%; Score 6; DB 2; Length 80;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKK 30  
|||||  
Db 41 KERKK 46

RESULT 37

B84145  
hypothetical protein BH3962 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 KKRERK 34  
|||||

Db 95 KKRERK 100

## RESULT 40

T07618

cold stress protein cill.8 - garden pea

C;Species: pisum sativum (garden pea)

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 24-Nov-1999

C;Accession: T07618

R;Kung, C.C.; Yeh, K.W.; Lin, C.Y.; Chen, Y.M.

Bot. Bull. Acad. Sin. 39, 9-15, 1998

A;Title: Characterization of a pea gene responsive to low temperature.

A;Reference number: Z16055

A;Accession: T07618

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-110 &lt;KUN&gt;

A;Cross-references: EMBL:U24398; NID:g2947080; PID:g2947081

A;Experimental source: cv. Taichung 9

C;Comment: This protein is cold-induced.

C;Superfamily: cold stress protein COR19

## Query Match

Best Local Similarity 6.2%; Score 6; DB 2; Length 110;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKRERK 29

|||||

Db 88 KKRERK 93

## RESULT 41

B40617

transcription repressor of ansAB operon ansR - Bacillus subtilis

C;Species: Bacillus subtilis

C;Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Jun-2000

C;Accession: B40617; B69586

R;Sun, D.; Setlow, P.

J. Bacteriol. 175, 2501-2506, 1993

A;Title: Cloning and nucleotide sequence of the Bacillus subtilis ansR gene, which encod

A;Reference number: A40617; MUID:93239674; PMID:8478318

A;Accession: B40617

A;Status: preliminary

A;Molecule type: nucleic acid

A;Residues: 1-116 &lt;SUN&gt;

A;Cross-references: GB:I08205; NID:g304138; PIDN:AAA72333.1; PID:g387575

A;Note: sequence extracted from NCBI backbone (NCBIN:130324, NCBIP:130334)

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: B69586

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-116 &lt;KUN&gt;

A;Cross-references: GB:Z99116; GB:AL009126; NID:g2634723; PIDN:CAB14291.1; PID:g2634794

A;Experimental source: strain 168

## C;Genetics:

A;Gene: ansR

C;Superfamily: probable transcription repressor yowR

C;Keywords: DNA binding; transcription regulation

## Query Match

Best Local Similarity 6.2%; Score 6; DB 2; Length 116;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TELRKK 25

|||||

Db 7 TELRKK 12

## RESULT 42

H72386

mazG-related protein - Thermotoga maritima (strain MSB8)

C;Species: Thermotoga maritima

C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C;Accession: H72386

R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; H

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson

C.M.

Nature 399, 323-329, 1999

A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genom

A;Reference number: A72200; MUID:99287316; PMID:10360571

A;Accession: H72386

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-116 &lt;ARN&gt;

A;Cross-references: GB:AE001716; GB:AE000512; NID:g4980853; PIDN:AAD35447.1; PID:g49

A;Experimental source: strain MSB8

C;Genetics:

A;Gene: TM0360

## Query Match

Best Local Similarity 6.2%; Score 6; DB 2; Length 116;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKK 30

|||||

Db 110 KERKKK 115

## RESULT 43

D72336

transcription regulator, GntR family - Thermotoga maritima (strain MSB8)

C;Species: Thermotoga maritima

C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C;Accession: D72336

R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; H

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,

C.M.

Nature 399, 323-329, 1999

A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A;Reference number: A72200; MUID:99287316; PMID:10360571

A;Accession: D72336

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-121 &lt;ARN&gt;

A;Cross-references: GB:AE001746; GB:AE000512; NID:g4981285; PIDN:AAD35848.1; PID:g498

A;Experimental source: strain MSB8

C;Genetics:

A;Gene: TM0766

C;Superfamily: transcription regulator gntR-related protein ytra

## Query Match

Best Local Similarity 6.2%; Score 6; DB 2; Length 121;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TELRKK 25

|||||

Db 90 TELRKK 95



Tue May 13 12:12:44 2003

RESULT 44  
F84868  
hypothetical protein At2g43640 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: F84868  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: F84868  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-121 <STO>  
A:Cross-references: GB:AE002093; NID:g2281106; PIDN:AAB64042.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g43640  
A:Map position: 2

Query Match	6.2%	Score 6;	DB 2;	Length 121;
Best Local Similarity	100.0%	Pred. No. 76;		
Matches 6;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	29	KKRERK 34
Db	94	KKRERK 99

RESULT 45  
T37059  
hypothetical protein SCJ21.10 - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T37059  
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1999  
A:Reference number: Z21620  
A:Accession: T37059  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-124 <SEE>  
A:Cross-references: EMBL:AL109747; PIDN:CAB52356.1; GSPDB:GN00070; SCOEDB:SCJ21.10  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SCJ21.10

Query Match	6.2%	Score 6;	DB 2;	Length 124;
Best Local Similarity	100.0%	Pred. No. 77;		
Matches 6;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	79	TGGCLP 84
Db	55	TGGCLP 60

Search completed: May 11, 2003, 20:13:40  
Job time : 34.0443 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:01:37 ; Search time 19.7434 Seconds  
(without alignments)  
203.775 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 97  
Sequence: 1 EVEVSRDHASLGSETLSQT.....LTGGCLPWATRSHLGRRKCS 97

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	8.2	67	1 RS21_AQUAE	O67028 aquifex aeo
2	7	7.2	129	1 RS9_CHLTR	O84128 chlamydia t
3	7	7.2	133	1 RS9_CHLMU	Q9pkr2 chlamydia m
4	7	7.2	134	1 RS9_CHLPN	Q928t8 chlamydia p
5	7	7.2	216	1 RGSJ_MOUSE	Q9cx84 mus musculus
6	7	7.2	310	1 YDHB_ECOLI	P37598 escherichia
7	7	7.2	356	1 YJN1_YEAST	P47015 saccharomyc
8	7	7.2	440	1 SOC4_HUMAN	Q8wxh5 homo sapien
9	7	7.2	583	1 CYSP_PLAVI	P42666 plasmodium
10	7	7.2	802	1 Y250_HUMAN	Q92540 homo sapien
11	7	7.2	858	1 41_MOUSE	P48193 mus musculu
12	7	7.2	864	1 41_HUMAN	P11171 homo sapien
13	7	7.2	1395	1 SP41_YEAST	P38904 saccharomyc
14	6	6.2	100	1 SMT3_ORYSA	P55857 oryza sativ
15	6	6.2	116	1 ANSR_BACSU	Q07683 bacillus su
16	6	6.2	121	1 SR14_ARATH	O04421 arabidopsis
17	6	6.2	129	1 RBFA_PSEAE	Q9hv56 pseudomonas
18	6	6.2	132	1 RBFA_VIBCH	Q9ku79 vibrio chol
19	6	6.2	132	1 SR14_ORYSA	O04433 oryza sativ
20	6	6.2	137	1 RR9_PORPU	P51291 porphyra pu
21	6	6.2	162	1 Y157_AQUAE	O66547 aquifex aeo
22	6	6.2	166	1 BAIE_EUBSP	P19412 eubacterium
23	6	6.2	184	1 TD52_HUMAN	P55327 homo sapien
24	6	6.2	184	1 TD52_RABIT	Q95212 oryctolagus
25	6	6.2	185	1 TD52_MOUSE	Q62393 mus musculu
26	6	6.2	188	1 P5PC_RABIT	P22398 oryctolagus
27	6	6.2	211	1 RCSA_ERWAM	P20098 erwinia amy
28	6	6.2	211	1 RCSA_ERWST	P27488 erwinia ste
29	6	6.2	231	1 NIA2_MAIZE	P39871 zea mays (m
30	6	6.2	233	1 MTMU_MYCSP	P43641 mycoplasma
31	6	6.2	234	1 A29B_DROSI	Q9u968 drosophila
32	6	6.2	237	1 BIOD_YEAST	P53630 saccharomyc
33	6	6.2	238	1 CW14_YEAST	O13547 saccharomyc

34	6	6.2	249	1 YCND_BACSU	P94424 bacillus su
35	6	6.2	252	1 AGL6_ARATH	P29386 arabidopsis
36	6	6.2	256	1 YQEU_BACSU	P54461 bacillus su
37	6	6.2	269	1 FPG_VIBCH	Q9kvc5 vibrio chol
38	6	6.2	279	1 NRTB_PHOLA	Q51881 phormidium
39	6	6.2	283	1 NADC_METJA	Q57916 methanococ
40	6	6.2	286	1 PARB_CHLPN	Q9z7m0 chlamydia p
41	6	6.2	313	1 EFTS_ANASP	Q8ymy3 anabaena sp
42	6	6.2	315	1 CPP1_ENTHI	Q01957 entamoeba h
43	6	6.2	316	1 L767_CAEEL	Q09517 caenorhabdi
44	6	6.2	319	1 H963_HUMAN	O14626 homo sapien
45	6	6.2	319	1 PRIM_BUCAP	P32000 buchnera ap
46	6	6.2	322	1 YMX7_CAEEL	P34515 caenorhabdi
47	6	6.2	339	1 RPOA_AEGTA	P92429 aegilops ta
48	6	6.2	339	1 RPOA_AGRCR	P92209 agropyron c
49	6	6.2	339	1 RPOA_AUSVE	P93993 australopyr
50	6	6.2	339	1 RPOA_BROIN	P92220 bromus iner
51	6	6.2	339	1 RPOA_CRIDE	P92225 crithopsis
52	6	6.2	339	1 RPOA_EREDI	P93974 eremopyrum
53	6	6.2	339	1 RPOA_FESFE	P93968 festucopsis
54	6	6.2	339	1 RPOA_FESSE	P93956 festucopsis
55	6	6.2	339	1 RPOA_HORVU	P92392 hordeum vul
56	6	6.2	339	1 RPOA_PSAFR	P92418 psathyrosta
57	6	6.2	339	1 RPOA_PSAFU	P93960 psathyrosta
58	6	6.2	339	1 RPOA_PSAST	P93962 psathyrosta
59	6	6.2	339	1 RPOA_SECT	P93964 secale stri
60	6	6.2	339	1 RPOA_THIBE	P92439 thinopyrum
61	6	6.2	339	1 RPOA_WHEAT	P12073 triticum ae
62	6	6.2	341	1 COA2_POVMK	P24596 mouse polyo
63	6	6.2	351	1 ADH_CLOBE	P25984 clostridium
64	6	6.2	357	1 GBA2_DICDI	P16051 dictyosteli
65	6	6.2	359	1 COBT_ECOLI	P36562 escherichia

ALIGNMENTS

RESULT 1	RS21_AQUAE	STANDARD;	PRT;	67 AA.
ID	RS21_AQUAE			
AC	O67028;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	30S ribosomal protein S21.			
GN	RPSU OR AQ_867 OR AQ_867A.			
OS	Aquifex aeolicus.			
OC	Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;			
OC	Aquifex.			
OX	NCBI_TaxID=63363;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=VF5;			
RX	MEDLINE=98196666; PubMed=9537320;			
RA	Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,			
RA	Graham D.E., Overbeek R., Snead M.A., Keller M., Aujaay M., Huber R.,			
RA	Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;			
RT	"The complete genome of the hyperthermophilic bacterium Aquifex			
RT	aeolicus.";			
RL	Nature 392:353-358(1998).			
CC	-1- SIMILARITY: BELONGS TO THE S21P FAMILY OF RIBOSOMAL PROTEINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AE000711; AAC06990.1; -.			
DR	InterPro; IPR001911; Ribosomal_S21.			
DR	Pfam; PF01165; Ribosomal_S21; 1.			

DR PRINTS; PR00976; RIBOSOMALS21.  
DR ProDom; PD005521; Ribosomal\_S21; 1.  
DR TIGRFAMs; TIGR00030; S21p; 1.  
DR PROSITE; PS01181; RIBOSOMAL\_S21; FALSE\_NEG.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 67 AA; 8261 MW; E1897087A487EF70 CRC64;  
  
Query Match 8.2%; Score 8; DB 1; Length 67;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 26 ERKKRER 33  
Db 43 ERKKRER 50

RESULT 2  
RS9\_CHLTR STANDARD; PRT; 129 AA.  
AC 084128;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 30S ribosomal protein S9.  
GN RPSI OR RS9 OR CT126.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D/UW-3/Cx;  
RX MEDLINE=99000809; PubMed=9784136;  
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
RA Davis R.W.;  
RT "Genome sequence of an obligate intracellular pathogen of humans:  
RT Chlamydia trachomatis.";  
RL Science 282:754-759(1998).  
CC -!- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE001286; AAC67717.1; -  
CC InterPro; IPR000754; Ribosomal\_S9.  
CC Pfam; PF00380; Ribosomal\_S9; 1.  
CC ProDom; PD001627; Ribosomal\_S9; 1.  
CC PROSITE; PS00360; RIBOSOMAL\_S9; FALSE\_NEG.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 129 AA; 14560 MW; 8AF05C6A824D2F8D CRC64;

Query Match 7.2%; Score 7; DB 1; Length 129;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 23 RKKERK 29  
Db 108 RKKERK 114  
  
RESULT 3  
RS9\_CHLMU STANDARD; PRT; 133 AA.  
AC Q9PKR2;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 30S ribosomal protein S9.

GN RPSI OR TC0402.  
OS Chlamydia muridarum.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83560;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Mopn / Nigg;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., Bass S.,  
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Dodson R.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
CC -!- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE002307; AAF39259.1; -  
CC TIGR; TC0402; -  
CC InterPro; IPR000754; Ribosomal\_S9.  
CC Pfam; PF00380; Ribosomal\_S9; 1.  
CC ProDom; PD001627; Ribosomal\_S9; 1.  
CC PROSITE; PS00360; RIBOSOMAL\_S9; FALSE\_NEG.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 133 AA; 14992 MW; DOE7BA313CBAAF35 CRC64;

Query Match 7.2%; Score 7; DB 1; Length 133;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 23 RKKERK 29  
Db 112 RKKERK 118

RESULT 4  
RS9\_CHLPN STANDARD; PRT; 134 AA.  
AC Q928T8; Q9JQJ9;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 30S ribosomal protein S9.  
GN RPSI OR RS9 OR CPN0246 OR CP0516.  
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CWL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RL Nat. Genet. 21:385-389(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;



RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
RT pneumoniae AR39.";  
RN Nucleic Acids Res. 28:1397-1406(2000).  
RP [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
RT from Japan and CWL029 from USA.";  
RL Nucleic Acids Res. 28:2311-2314(2000).  
CC -1- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE001610; AAD18399.1; -  
CC EMBL; AE002211; AAF38343.1; -  
CC EMBL; AP002545; BAA98456.1; -  
CC TIGR; CP0516;  
CC InterPro; IPR000754; Ribosomal\_S9.  
CC Pfam; PF00380; Ribosomal\_S9; 1.  
CC ProDom; PD001627; Ribosomal\_S9; 1.  
CC PROSITE; PS00360; RIBOSOMAL\_S9; 1.  
CC Ribosomal protein; Complete proteome.  
KW SEQUENCE 134 AA; 15195 MW; A9869207154F4AE8 CRC64;  
SQ

Query Match 7.2%; Score 7; DB 1; Length 134;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERK 29  
Db 113 RKKERK 119

RESULT 5  
RGSJ\_MOUSE STANDARD; PRT; 216 AA.  
ID RGSJ\_MOUSE Q99L50;  
AC Q9CX84; Q99L50;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Regulator of G-protein signaling 19 (RGS19).  
GN RGS19.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast tumor;  
RA Strausberg R.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: INHIBITS SIGNAL TRANSDUCTION BY INCREASING THE GTPASE  
CC ACTIVITY OF G PROTEIN ALPHA SUBUNITS THEREBY DRIVING THEM INTO  
CC THEIR INACTIVE GDP-BOUND FORM. BINDS TO G-ALPHA SUBFAMILY 1  
CC MEMBERS, WITH THE ORDER G(I)A3 > G(I)A1 > G(O)A >> G(Z)A/G(I)A2  
CC ACTIVITY ON G(Z)-ALPHA IS INHIBITED BY PHOSPHORYLATION AND  
CC PALMITOYLATION OF THE G-PROTEIN (BY SIMILARITY).  
CC -1- SUBUNIT: INTERACTS WITH GIPC PDZ DOMAIN (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Membrane-bound (By similarity).  
CC -1- PTM: FATTY ACYLATED. HEAVILY PALMITOYLATED IN THE CYSTEINE STRING  
CC MOTIF (BY SIMILARITY).  
CC -1- PTM: PHOSPHORYLATED, MAINLY ON SERINE RESIDUES (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 1 RGS DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AK019401; BAB31703.1; -  
CC EMBL; BC003838; AAH03838.1; -  
CC HSSP; P49795; 1CMZ.  
CC MGD; MGI:1915153; Rgs19.  
CC InterPro; IPR000342; Regl\_Gprotein.  
CC Pfam; PF00615; RGS; 1.  
CC PRINTS; PR01301; RGS-PROTEIN.  
CC ProDom; PD001580; Reg\_of\_prg; 1.  
CC SMART; SM00315; RGS; 1.  
CC PROSITE; PS0132; RGS; 1.  
KW Signal transduction inhibitor; Membrane; Lipoprotein; Palmitate;  
KW Phosphorylation.  
FT DOMAIN 39 49 POLY-CYS.  
FT DOMAIN 90 206 RGS.  
FT DOMAIN 207 216 INTERACTS WITH GIPC (BY SIMILARITY).  
FT MOD\_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 151 151 PHOSPHORYLATION (BY MAPK1 AND MAPK3) (BY  
FT SIMILARITY).  
FT CONFLICT 82 82 K -> E (IN REF. 2).  
FT SEQUENCE 216 AA; 24677 MW; 4F166A6607184F31 CRC64;  
SQ

Query Match 7.2%; Score 7; DB 1; Length 216;  
Best Local Similarity 100.0%; Pred. No. 5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 PPSPEV 72  
Db 78 PPSPEV 84

RESULT 6  
YDHB\_ECOLI STANDARD; PRT; 310 AA.  
ID YDHB\_ECOLI P77677;  
AC P37598; P77677;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical transcriptional regulator ydhB.  
GN YDHB OR B1659 OR Z2682 OR ECS2368.  
OS Escherichia coli, and  
OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562, 83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / RR28;  
RA Eberhardt S.M.R., Richter G., Gimbel W., Werner T., Bacher A.;  
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MGI655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,  
RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,  
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 28.0-40.1 min region on the linkage map.";  
RL DNA Res. 3:363-377(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grothbeck E.J., Davis N.W., Llm A., Dimallanta E.T., Potamouis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
RL Nature 409:529-533(2001).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shingagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22(2001).  
RN [6]  
RP SEQUENCE OF 106-310 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=90126847; PubMed=2404765;  
RA Meng L.M., Kilstrup M., Nygaard P.;  
RT "Autoregulation of PurR repressor synthesis and involvement of purR  
RT in the regulation of purB, purC, purL, purM and guaBA expression in  
RT Escherichia coli.";  
RL Eur. J. Biochem. 187:373-379(1990).  
RN [7]  
RP IDENTIFICATION.  
RA Rudd K.E.;  
RL Unpublished observations (FEB-1994).  
CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X69109; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; AE000261; AAC74731.1; -.  
DR EMBL; D90809; BAA15425.1; -.  
DR EMBL; D90810; BAA15434.1; -.  
DR EMBL; AE005389; AAG56648.1; -.  
DR EMBL; AP002558; BAB35791.1; -.  
DR EMBL; X51368; -; NOT\_ANNOTATED\_CDS.  
DR EcoGene; EGI2140; ydhB.  
DR InterPro; IPR000847; HTH\_LysR.  
DR InterPro; IPR005119; LysR\_subst.  
DR Pfam; PF00126; HTH\_1; 1.  
DR Pfam; PF03466; LysR\_substrate; 1.  
DR PRINTS; PR00039; HTHLYSR.  
DR PROSITE; PS00044; HTH\_LYSR\_FAMILY; 1.  
KW Hypothetical protein; Transcription regulation; DNA-binding;  
KW Complete proteome.  
FT DNA\_BIND 19 38 H-T-H MOTIF (BY SIMILARITY).  
FT CONFLICT 253 253 K -> E (IN REF. 3).  
SQ SEQUENCE 310 AA; 35250 MW; BBC7A2F2B2AC351A CRC64;  
Query Match 7.2%; Score 7; DB 1; Length 310;  
Best Local Similarity 100.0%; Pred. No. 6.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 11 LGDSETL 17  
Db 288 LGDSETL 294  
RESULT 7  
YJNL\_YEAST  
ID YJNL\_YEAST STANDARD; PRT; 356 AA.  
AC P47015;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Hypothetical 41.5 kDa protein in MRS3-URA2 intergenic region.  
GN YJL131C OR J0682.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / FY1679;  
RX MEDLINE=96408771; PubMed=8813765;  
RA Katsoulou C., Tzermia M., Tavernarakis N., Alexandraki D.;  
RT "Sequence analysis of a 40.7 kb segment from the left arm of yeast  
RT chromosome X reveals 14 known genes and 13 new open reading frames  
RT including homologues of genes clustered on the right arm of  
RT chromosome XI.";  
RL Yeast 12:787-797(1996).  
RN [2]  
RP This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X87371; CAA60824.1; -.  
DR EMBL; Z49406; CAA89426.1; -.  
DR SGD; S0003667; YJL131C.  
KW Hypothetical protein.  
SQ SEQUENCE 356 AA; 41461 MW; BCF907223417B6C2 CRC64;  
Query Match 7.2%; Score 7; DB 1; Length 356;

```

Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      21  ELKKER 27
          |||||
Db      334  ELKKER 340

RESULT 8
SOC4_HUMAN
ID      SOC4_HUMAN      STANDARD;      PRT;      440 AA.
AC      Q8WXH5;
DT      15-JUN-2002 (Rel. 41, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      suppressor of cytokine signaling 4 (SOCS-4).

```

15-JUN-2002 (Rel. 41, Last sequence update)  
DT DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Suppressor of cytokine signaling 4 (SOCS-4).  
GN SOCS4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hilton D.J., Alexander W.S., Nicola N.A.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: SOCS family proteins form part of a classical negative  
CC feedback system that regulates cytokine signal transduction.  
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SOCS BOX DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	InterPro: IPR001496; SOCS.
DR	PROSITE: PS50001; SH2; 1.
DR	PROSITE: PS50225; SOCS; 1.
KW	SH2 domain; Growth regulation; Signal transduction inhibitor.
FT	DOMAIN 286 381 SH2.
FT	DOMAIN 376 425 SOCS BOX.
SC	SEQUENCE 440 AA; 50623 MW; A4A747AB7F6FF08C CRC64;

Query Match	7.2%	Score 7;	DB 1;	Length 440;
Best Local Similarity	100.0%	Pred. No. 9.4;		
Matches	7;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
			0;	
QY	17	LSQTELR	23	
Db	166	LSQTELR	172	

CYSP_PLAVI	STANDARD;	PRT;	583 AA.
ID	CYSP_PLAVI		
AC	P42666;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Cysteine proteinase precursor (EC 3.4.22.-).		
OS	Plasmodium vivax.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5855;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Salvador I;		
RX	MEDLINE=94334995; PubMed=8057374;		
RA	Rosenthal P.J.; Ring C.S.; Chen X.; Cohen F.E.;		



```
RL DNA Res. 3:321-329(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D87437; BAA1381.1; -.
DR Genew; HGNC:16792; Clorf16.
KW Hypothetical protein.
SQ SEQUENCE 802 AA; 88998 MW; E9AF1C0AA1790023 CRC64;

Query Match
Best Local Similarity 7.2%; Score 7; DB 1; Length 802;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 SQTELK 24
Db 282 SQTELK 288

RESULT 11
41_MOUSE
ID 41_MOUSE STANDARD; PRT; 858 AA.
AC P48193;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein 4.1 (Band 4.1) (P4.1) (4.1R).
GN EPB41 OR EPB4.1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=BA1B/C;
RX MEDLINE=93155238; PubMed=8429050;
RA Huang J.-P., Tang C.-J.C., Kou G.-H., Marchesi V.T., Benz E.J. Jr.,
RA Tang T.K.;
RT "Genomic structure of the locus encoding protein 4.1. Structural
RT basis for complex combinatorial patterns of tissue-specific
RT alternative RNA splicing.";
RL J. Biol. Chem. 268:3758-3766(1993).
RN [2]
RP SEQUENCE OF 709-713, AND CHARACTERIZATION OF CARBOXY-TERMINAL DOMAIN.
RX MEDLINE=21325946; PubMed=11432737;
RA Scott C., Phillips G.W., Baines A.J.;
RA "Properties of the C-terminal domain of 4.1 proteins.";
RL Eur. J. Biochem. 268:3709-3717(2001).
CC -!- FUNCTION: Protein 4.1 is a major structural element of the
CC erythrocyte membrane skeleton. It plays a key role in regulating
CC membrane physical properties of mechanical stability and
CC deformability by stabilizing spectrin-actin interaction. Binds
CC with a high affinity to glycophorin and with lower affinity to
CC band III protein.
CC -!- ALTERNATIVE PRODUCTS: A number of isoforms are produced by
CC alternative splicing.
CC -!- MASS SPECTROMETRY: MW=17199.3; METHOD=Electrospray; RANGE=709-858.
CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L00919; AAA37123.1; -.

```

```
DR MGD; MGI:95401; Epb4.1.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS50057; BAND_41_3; 1.
KW Structural protein; Alternative splicing; Cytoskeleton;
KW Phosphorylation.
FT DOMAIN 208 422 BAND 4.1-LIKE.
FT DOMAIN 489 608 HYDROPHILIC.
FT DOMAIN 609 707 SPECTRIN-ACTIN-BINDING.
FT DOMAIN 710 858 CARBOXYL-TERMINAL (CTD).
SQ SEQUENCE 858 AA; 95990 MW; 5F2FEF077946134E CRC64;

Query Match
Best Local Similarity 7.2%; Score 7; DB 1; Length 858;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKRR 33
Db 641 RKKRR 647

RESULT 12
41_HUMAN
ID 41_HUMAN STANDARD; PRT; 864 AA.
AC P11171; P11176; Q9Y578; Q9Y579; Q14245;
DT 01-JUL-1989 (Rel. 11, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein 4.1 (Band 4.1) (P4.1) (EPB4.1) (4.1R).
GN EPB41 OR E41P.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (NON-ERYTHROID FORM).
RX MEDLINE=88234496; PubMed=3375238;
RA Tang T.K., Leto T.L., Correas I., Alonso M.A., Marchesi V.T.,
RA Benz E.J. Jr.;
RT "Selective expression of an erythroid-specific isoform of protein
RT 4.1.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:3713-3717(1988).
RN [2]
RP SEQUENCE FROM N.A. (NON-ERYTHROID ISOFORM).
RX MEDLINE=89132003; PubMed=3223413;
RA Tang T.K., Leto T.L., Marchesi V.T., Benz E.J. Jr.;
RT "Expression of specific isoforms of protein 4.1 in erythroid and non-
RT erythroid tissues.";
RL Adv. Exp. Med. Biol. 241:81-95(1988).
RN [3]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 378-393 (ERYTHROID ISOFORM).
RC TISSUE=Reticulocytes;
RX MEDLINE=87092279; PubMed=3467321;
RA Conboy J.G., Kan Y.W., Shohet S.B., Mohandas N.;
RT "Molecular cloning of protein 4.1, a major structural element of the
RT human erythrocyte membrane skeleton.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9512-9516(1986).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RX MEDLINE=91217063; PubMed=2022644;
RA Conboy J.G., Chan J.Y.C., Chasis J.A., Kan Y.W., Mohandas N.;
RT "Tissue- and development-specific alternative RNA splicing regulates
RT expression of multiple isoforms of erythroid membrane protein 4.1.";
RL J. Biol. Chem. 266:8273-8280(1991).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Huang S.C., Wang C., Lichtenauer U., Vortmeyer A., Zhuang Z.;
RT "Sequence of protein 4.1 from a human neuroblastoma cell line: LAN5.";
```



DR InterPro; IPR000299; Band\_4.1.  
DR Pfam; PF00373; Band\_4.1; 1.  
DR PRINTS; PRO0935; BAND41.  
DR SMART; SM00295; B41; 1.  
DR PROSITE; PS00660; BAND\_4.1\_1; 1.  
DR PROSITE; PS00661; BAND\_4.1\_2; 1.  
DR PROSITE; PS00662; BAND\_4.1\_3; 1.  
DR Structural protein; Alternative splicing; Cytoskeleton; Elliptocytosis;  
KW Phosphorylation; Pyropoikilocytosis; Glycoprotein; Polymorphism;  
KW Hereditary hemolytic anemia; Polymorphism;  
FT DOMAIN 207 421  
FT DOMAIN 488 614  
FT DOMAIN 615 713  
FT DOMAIN 714 864  
FT MOD\_RES 660 660  
FT VARSPLIC 1 209  
FT VARSPLIC 228 262  
FT VARSPLIC 616 648  
FT VARSPLIC 635 648  
FT VARSPLIC 649 669  
FT VARSPLIC 772 805  
FT VARIANT 214 214  
FT CONFLICT 51 51  
FT CONFLICT 76 76  
FT CONFLICT 168 168  
FT CONFLICT 259 259  
FT CONFLICT 665 665  
FT CONFLICT 679 679  
FT CONFLICT 802 802  
SQ SEQUENCE 864 AA; 97016 MW; B4731249D7FBE31 CRC64;  
Query Match 7.2%; Score 7; DB 1; Length 864;  
Best Local Similarity 100.0%; Pred. No. 17; 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;  
QY 27 RKKRRER 33  
Db 647 RKKRRER 653  
RESULT 13  
SP41\_YEAST  
ID SP41\_YEAST STANDARD; PRT; 1395 AA.  
AC P38904;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE SPP41 protein.  
GN SPP41 OR YDR464W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A364A;  
RX MEDLINE=94274035; PubMed=8005438;  
RA Madgock J.R.; Weidenhammer E.M.; Adams C.C.; Lunz R.L.;  
RA Woolford J.L. Jr.;  
RT "Extragenic suppressors of Saccharomyces cerevisiae prp4 mutations  
RT identify a negative regulator of prp genes.";  
RL Genet. 136:833-847(1994).  
CC -!- FUNCTION: NEGATIVE REGULATOR OF PRP3 AND PRP4 GENES. INTERACTS  
CC WITH PRP8 AND RAP1.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
DR Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
[6] SEQUENCE OF 157-227 FROM N.A., AND VARIANT ILE-214.  
RA Lichtenauer U., Huang S.C., Vortmeyer A., Zhuang Z.;  
RT "Valine to isoleucine polymorphism in exon 4 of human protein 4.1.";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
[7] SEQUENCE OF 648-714.  
RP MEDLINE=87008553; PubMed=3531202;  
RX Correia I., Speicher D.W., Marchesi V.T.;  
RA "Structure of the spectrin-actin binding site of erythrocyte protein  
RT 4.1.";  
RL J. Biol. Chem. 261:13362-13366(1986).  
[8] CHARACTERIZATION OF CARBOXY-TERMINAL DOMAIN.  
RP MEDLINE=21325946; PubMed=11432737;  
RX Scott C., Phillips G.W., Baines A.J.;  
RA "Properties of the C-terminal domain of 4.1 proteins.";  
RT Eur. J. Biochem. 268:3709-3717(2001).  
[9] PHOSPHORYLATION OF TYR-660.  
RP MEDLINE=91271361; PubMed=1647028;  
RX Subrahmanyam G., Bertics P.J., Anderson R.A.;  
RA "Phosphorylation of protein 4.1 on tyrosine-418 modulates its  
RT function in vitro.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:5222-5226(1991).  
[10] STRUCTURE OF CARBOHYDRATES.  
RP MEDLINE=90036892; PubMed=2808371;  
RX Inaba M., Maede Y.;  
RA "O-N-acetyl-D-glucosamine moiety on discrete peptide of multiple  
RT protein 4.1 isoforms regulated by alternative pathways.";  
RL J. Biol. Chem. 264:18149-18155(1989).  
CC -!- FUNCTION: Protein 4.1 is a major structural element of the  
CC erythrocyte membrane skeleton. It plays a key role in regulating  
CC membrane physical properties of mechanical stability and  
CC deformability by stabilizing spectrin-actin interaction. Binds  
CC with a high affinity to glyophorin and with lower affinity to  
CC band III protein. Associates with the nuclear mitotic apparatus.  
CC -!- ALTERNATIVE PRODUCTS: 6 isoforms; 1 (shown here), 2, 3, erythroid,  
CC non-erythroid A and non-erythroid B; are produced by alternative  
CC splicing.  
CC -!- PTM: Phosphorylated at multiple sites by different protein kinases  
CC and each phosphorylation event selectively modulates the protein's  
CC functions.  
CC -!- PTM: Phosphorylation on Tyr-660 reduces the ability of 4.1 to  
CC promote the assembly of the spectrin/actin/4.1 ternary complex.  
CC -!- PTM: O-glycosylated; contains N-acetylglucosamine side chains  
CC in the C-terminal domain.  
CC -!- DISEASE: Defects in EPB41 are a cause of hereditary elliptocytosis  
CC (HE) and of hereditary pyropoikilocytosis (HPP).  
CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; J03796; AAA35793.1; -.  
EMBL; J03796; AAA35794.1; -.  
EMBL; M14993; AAA35795.1; -.  
EMBL; M61733; AAA35797.1; -.  
EMBL; AF156225; AAD42222.1; -.  
EMBL; AF156226; AAD42223.1; -.  
PIR; A60244; MMHUL4.  
PIR; A26656; MMHUE4.  
DR GlycoSuiteDB; P11171; -.  
DR Genew; HGNC:3377; EPB41.  
DR MIM; 130500; -.  
DR MIM; 266140; -.

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; U03673; AAA20494.1; -  
DR PIR; S47864; S47864.  
DR SGD; S0002872; SPP41.  
DR InterPro; IPR003903; UIM.  
KW Nuclear protein.  
FT DOMAIN 683  
SQ SEQUENCE 1395 AA; 156933 MW; 6EC9DA9BB21B3471 CRC64;  
NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

Query Match  
Best Local Similarity 7.2%; Score 7; DB 1; Length 1395;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 24 KKERKKK 30  
Db 695 KKERKKK 701  
|||||

RESULT 14  
SMT3\_ORYSA  
ID SMT3\_ORYSA STANDARD; PRT; 100 AA.  
AC P55857;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Ubiquitin-like protein SMT3.  
GN SMT3.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97237059; PubMed=9119407;  
RA Lapenta V., Chiurazzi P., van der Spek P.J., Pizzuti A.,  
RA Hanaoka F., Brahe C.;  
RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to  
RT chromosome 21qter and defines a novel gene family.";  
RL Genomics 40:362-367(1997).  
CC -!- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.  
-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; X99608; CAA67922.1; -  
DR HSSP; Q93068; 1A5R.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF00240; ubiquitin; 1.  
DR SMART; SM00213; UBQ; 1.  
DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
FT DOMAIN 19 96 UBIQUITIN-LIKE.  
SQ SEQUENCE 100 AA; 10928 MW; 624E44BC94C00268 CRC64;  
NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

Query Match  
Best Local Similarity 6.2%; Score 6; DB 1; Length 100;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 79 TGGCLP 84  
Db 94 TGGCLP 99  
|||||

RESULT 15  
ANSR\_BACSU  
ID ANSR\_BACSU STANDARD; PRT; 116 AA.  
AC Q07683;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ans operon repressor protein.  
GN ANSR.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93239674; PubMed=8478318;  
RA Sun D., Setlow P.;  
RT "Cloning and nucleotide sequence of the Bacillus subtilis ansR gene,  
RT which encodes a repressor of the ans operon coding for L-asparaginase  
RT and L-aspartase.";  
RL J. Bacteriol. 175:2501-2506(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168 / JH642;  
RX MEDLINE=97124195; PubMed=8969508;  
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,  
RA Kobayashi Y.;  
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of  
RT the Bacillus subtilis genome containing the skin element and many  
RT sporulation genes.";  
RL Microbiology 142:3103-3111(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE ANS OPERON CODING FOR  
CC L-ASPARAGINASE AND L-ASPARTASE. NH4 MAY INFLUENCE THIS REPRESSION.  
CC -!- SIMILARITY: BELONGS TO THE PBSX(XRE) FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; L08205; AAA72333.1; -.  
DR EMBL; D84432; BAA12641.1; -.  
DR EMBL; Z99116; CAB14291.1; -.  
DR PIR; B40617; B40617.  
DR Subtilist; BGI0299; ansR.  
DR InterPro; IPR001387; HTH\_3.  
DR Pfam; PF01381; HTH\_3; 1.  
DR SMART; SM00530; HTH\_XRE; 1.  
DR Transcription regulation; Repressor; DNA-binding; Complete proteome.  
KW DNA\_BIND 17 36 H-T-H MOTIF (BY SIMILARITY).  
FT SEQUENCE 116 AA; 13231 MW; 29B86500ACB901F8 CRC64;  
SQ

Query Match 6.2%; Score 6; DB 1; Length 116;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TELKK 25  
|||||  
Db 7 TELKK 12

RESULT 16  
SR14\_ARATH STANDARD; PRT; 121 AA.

AC O04421; O22839;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Signal recognition particle 14 kDa protein (SRP14).  
GN SRP14 OR AT2G43640 OR F18O19.25.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia; TISSUE=Root;  
RA Bui N., Wolff N., Strub K.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=20083487; PubMed=10617197;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-J., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,  
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,  
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,  
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,  
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,  
RA Venter J.C.;  
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
thaliana.";  
RL Nature 402:761-768(1999).

CC -!- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE  
CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC RETICULUM  
CC MEMBRANE. SRP9 TOGETHER WITH SRP14 AND THE ALU PORTION OF THE SRP  
CC RNA, CONSTITUTES THE ELONGATION ARREST DOMAIN OF SRP. THE COMPLEX  
CC OF SRP9 AND SRP14 IS REQUIRED FOR SRP RNA BINDING (BY SIMILARITY).  
CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE  
CC OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,  
CC SRP19, SRP14 AND SRP9 (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: BELONGS TO THE SRP14 FAMILY.  
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; Y10116; CAA71202.1; -.  
DR EMBL; AC002333; AAB64042.1; -.  
DR HSP; P16254; 1914.  
DR InterPro; IPR003210; SRP14.  
DR Pfam; PF02290; SRP14; 1.  
DR Signal recognition particle; RNA-binding.  
KW CONFLICT 116 116 P -> T (IN REF. 2).  
FT SEQUENCE 121 AA; 13777 MW; 216D2AA83B24E7DD CRC64;  
SQ

Query Match 6.2%; Score 6; DB 1; Length 121;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 KKRERK 34  
|||||  
Db 94 KKRERK 99

RESULT 17

RBFA\_PSEAE STANDARD; PRT; 129 AA.

AC Q9HV56;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ribosome-binding factor A.  
GN RBFA OR PA4743.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
CC -!- FUNCTION: Associates with free 30S ribosomal subunits (but not  
CC with 30S subunits that are part of 70S ribosomes or polysomes).  
CC Essential for efficient processing of 16S rRNA. May interact with  
CC the 5'terminal helix region of 16S Rrna (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -!- SIMILARITY: BELONGS TO THE RBFA FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AE004888; AAG08129.1; -.  
DR InterPro; IPR000238; Rib\_bind\_facta.  
DR Pfam; PF02033; RBFA; 1.  
DR ProDom; PD007327; Rib\_bind\_facta; 1.  
DR TIGRFAMS; TIGR00082; rbfa; 1.  
DR PROSITE; PS01319; RBFA; 1.  
KW rRNA processing; Complete proteome.



```
SQ SEQUENCE 129 AA; 14547 MW; 1BE5E7FFC08FA9D6 CRC64;
Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 129;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VEVS RD 7
    |||||
DB 40 VEVS RD 45

RESULT 18
RBFA_VIBCH
ID RBFA_VIBCH STANDARD; PRT; 132 AA.
AC Q9KU79;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribosome-binding factor A.
GN RBFA OR VC0644.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
CC -!- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC with 30S subunits that are part of 70S ribosomes or polysomes).
CC Essential for efficient processing of 16S rRNA. May interact with
CC the 5'terminal helix region of 16S rRNA (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE RBFA FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE004150; AAF93810.1; ALT_INIT.
TIGR; VC0644;
DR InterPro; IPR000238; Rib_bind_factA.
DR Pfam; PF02033; RBFA; 1.
DR ProDom; PD007327; Rib_bind_factA; 1.
DR TIGRFAMs; TIGR00082; rbfa; 1.
DR PROSITE; PS01319; RBFA; 1.
KW rRNA processing; Complete proteome.
SQ SEQUENCE 132 AA; 15404 MW; 26D950866B1C6BA0 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 132;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VEVS RD 7
    |||||
DB 40 VEVS RD 45

RESULT 19
SR14_ORYSA
```

```
ID SR14_ORYSA STANDARD; PRT; 132 AA.
AC O04433;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Signal recognition particle 14 kDa protein (SRP14).
GN SRP14.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare; TISSUE=Etolated root;
RA Bui N., Wolff N., Strub K.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE
CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC RETICULUM
CC MEMBRANE. SRP9 TOGETHER WITH SRP14 AND THE ALU PORTION OF THE SRP
CC RNA, CONSTITUTES THE ELONGATION ARREST DOMAIN OF SRP. THE COMPLEX
CC OF SRP9 AND SRP14 IS REQUIRED FOR SRP RNA BINDING (BY SIMILARITY).
CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE
CC OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,
CC SRP19, SRP14 AND SRP9 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE SRP14 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Y10118; CAA71204.1;
DR HSP; P16254; 1914.
DR InterPro; IPR003210; SRP14.
DR Pfam; PF02290; SRP14; 1.
KW Signal recognition particle; RNA-binding.
SQ SEQUENCE 132 AA; 14838 MW; 374DA93020890A73 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 132;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 KKRRK 34
    |||||
DB 94 KKRRK 99

RESULT 20
RR9_PORPU
ID RR9_PORPU STANDARD; PRT; 137 AA.
AC P51291;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Chloroplast 30S ribosomal protein S9.
GN RPS9.
OS Porphyra purpurea.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Avonport;
RA Reith M.E., Munnholland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT genome.";
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC -!- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.
```



-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; U38804; AAC08177.1; -.  
InterPro; IPR000754; Ribosomal\_S9.  
Pfam; PF00380; Ribosomal\_S9; 1.  
ProDom; PD001627; Ribosomal\_S9; 1.  
PROSITE; PS00360; RIBOSOMAL\_S9; 1.  
Ribosomal protein; Chloroplast.  
SEQUENCE 137 AA; 15033 MW; AA6B96DFAE728A57 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 137;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 GRRKCS 97  
| | | | |  
Db 16 GRRKCS 21

RESULT 21  
Y157\_AQUAE STANDARD; PRT; 162 AA.  
AC O66547;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein AQ\_157 precursor.  
GN AQ\_157.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;  
OC Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
aeolicus.";  
RL Nature 392:353-358(1998).  
CC -1- SIMILARITY: BELONGS TO THE OMPH/HLPA FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; AE000676; AAC06512.1; -.  
Hypothetical protein; Coiled coil; Signal; Complete proteome.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 162 HYPOTHETICAL PROTEIN AQ\_157.  
FT DOMAIN 32 118 COILED COIL (POTENTIAL).  
FT SEQUENCE 162 AA; 18878 MW; 61DB961E197471A6 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 162;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 QTELK 24  
| | | | |  
Db 43 QTELK 48

RESULT 22  
BAIE\_EUBSP STANDARD; PRT; 166 AA.  
ID BAIE\_EUBSP  
AC P19412;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Bile acid-inducible operon protein E.  
GN BAIE.  
OS Eubacterium sp. (strain VPI 12708).  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=29347;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-31.  
RX MEDLINE=91072253; PubMed=2254270;  
RA Mallonee D.H., White W.B., Hylemon P.B.;  
RT "Cloning and sequencing of a bile acid-inducible operon from  
Eubacterium sp. strain VPI 12708.";  
RL J. Bacteriol. 172:7011-7019(1990).  
CC -1- PATHWAY: Bile acid catabolism.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; U57489; AAC45413.1; -.  
PIR; D37844; D37844.  
KW Bile acid catabolism.  
FT CONFLICT 2 2 T -> F (IN REF. 1; AA SEQUENCE).  
SQ SEQUENCE 166 AA; 19533 MW; 1CBCE86C85ADC3E5 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 166;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 SPKEVT 73  
| | | | |  
Db 63 SPKEVT 68

RESULT 23  
TD52\_HUMAN STANDARD; PRT; 184 AA.  
ID TD52\_HUMAN  
AC P55327; Q13056;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Tumor protein D52 (N8 protein).  
GN TPDS2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Breast carcinoma;  
MEDLINE=95316866; PubMed=7796418;  
RA Byrne J.A., Tomasetto C., Garnier J.M., Rouyer N., Mattel M.-G.,  
RA Bellocq J.P., Rio M.C., Basset P.;  
RT "A screening method to identify genes commonly overexpressed in  
carcinomas and the identification of a novel complementary DNA  
sequence.";  
RL Cancer Res. 55:2896-2903(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96197754; PubMed=8632896;

RA Chen S.L., Maroulakou I.G., Green J.E., Romano-Spica V., Modi W.,  
RA Lautenberger J., Bhat N.K.;  
RT "Isolation and characterization of a novel gene expressed in multiple  
RT cancers";  
RL Oncogene 12:741-751(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP INTERACTIONS.  
RX MEDLINE=98143307; PubMed=9484778;  
RA Byrne J.A., Nourse C.R., Basset P., Gunning P.;  
RT "Identification of homo- and heteromeric interactions between members  
RT of the breast carcinoma-associated D52 protein family using the yeast  
RT two-hybrid system.";  
RL Oncogene 16:873-881(1998).  
CC -1- SUBUNIT: FORMS HOMODIMER OR HETERODIMER WITH OTHER MEMBERS OF THE  
CC FAMILY.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN CANCER CELLS.  
CC -1- SIMILARITY: BELONGS TO THE TPD52 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U18914; AAC50183.1; -.  
DR EMBL; S82081; AAB36475.1; -.  
DR EMBL; BC018117; AAB18117.1; -.  
DR Genew; HGNC:12005; TPD52.  
DR MIM; 604068; -.  
KW Coiled coil.  
FT DOMAIN 22 74 COILED COIL (POTENTIAL).  
SQ SEQUENCE 184 AA; 19863 MW; 4821EC86D1C3339D CRC64;  
  
Query Match 6.2%; Score 6; DB 1; Length 184;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 14 SETLSQ 19  
Db 100 SETLSQ 105  
  
RESULT 24  
TD52\_RABIT  
ID TD52\_RABIT STANDARD; PRT; 184 AA.  
AC Q95212;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Tumor protein D52 (28 kDa calcium-dependent phosphoprotein) (pp28).  
GN TPD52 OR CSPP28.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Gastric fundus;  
RX MEDLINE=96355320; PubMed=8702730;  
RA Parente J.A., Goldenring J.R., Petropoulos A.C., Hellman U.,  
RA Chew C.S.;  
RT "Purification, cloning, and expression of a novel, endogenous,  
RT calcium-sensitive, 28-kDa phosphoprotein.";  
RL J. Biol. Chem. 271:20096-20101(1996).  
CC -1- SUBUNIT: FORMS HOMODIMER OR HETERODIMER WITH OTHER MEMBERS OF THE  
CC FAMILY (BY SIMILARITY).  
-----

CC -1- PTM: PHOSPHORYLATED IN A CALCIUM/CALMODULIN-DEPENDENT MANNER.  
CC -1- SIMILARITY: BELONGS TO THE TPD52 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U35428; AAC48616.1; -.  
KW Coiled coil; Phosphorylation.  
FT DOMAIN 22 74 COILED COIL (POTENTIAL).  
SQ SEQUENCE 184 AA; 19809 MW; 8CEA3C2CD6AC3DC4 CRC64;  
  
Query Match 6.2%; Score 6; DB 1; Length 184;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 14 SETLSQ 19  
Db 100 SETLSQ 105  
  
RESULT 25  
TD52\_MOUSE  
ID TD52\_MOUSE STANDARD; PRT; 185 AA.  
AC Q62393;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Tumor protein D52 (mD52).  
GN TPD52.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Mammary gland;  
RX MEDLINE=97001154; PubMed=8812487;  
RA Byrne J.A., Mattei M.-G., Basset P.;  
RT "Definition of the tumor protein D52 (TPD52) gene family through  
RT cloning of D52 homologues in human (hD53) and mouse (mD52).";  
RL Genomics 35:523-532(1996).  
CC -1- SUBUNIT: FORMS HOMODIMER OR HETERODIMER WITH OTHER MEMBERS OF THE  
CC FAMILY (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE TPD52 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U44426; AAB40897.1; -.  
DR MGD; MGI:107749; Tpd52.  
KW Coiled coil.  
FT DOMAIN 22 74 COILED COIL (POTENTIAL).  
SQ SEQUENCE 185 AA; 20059 MW; 345B487842135D33 CRC64;  
  
Query Match 6.2%; Score 6; DB 1; Length 185;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 14 SETLSQ 19  
Db 100 SETLSQ 105



DR ProDom; PD000307; HTH\_LuxR; 1.  
DR SMART; SM00421; HTH\_LuxR; 1.  
DR PROSITE; PS00622; HTH\_LuxR\_FAMILY; 1.  
KW Transcription regulation; DNA-binding; Activator.  
FT DNA\_BIND 159 178 H-T-H MOTIF (BY SIMILARITY).  
SQ SEQUENCE 211 AA; 24349 MW; 51BC162269735C55 CRC64;

Query Match  
Best Local Similarity 6.2%; Score 6; DB 1; Length 211;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TLSQTE 21  
Db 141 TLSQTE 146

RESULT 28  
RCSA\_ERWST  
ID RCSA\_ERWST STANDARD; PRT; 211 AA.  
AC P27488;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Colanic acid capsular biosynthesis activation protein A.  
GN RCSA.  
OS Erwinia stewartii.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Pantoea.  
OX NCBI\_TaxID=66271;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SS104;  
RX MEDLINE=91375445; PubMed=1896018;  
RA Poetter K.; Coplin D.L.;  
RT "Structural and functional analysis of the rcsA gene from Erwinia stewartii."  
RL Mol. Gen. Genet. 229:155-160(1991).  
CC -!- FUNCTION: POSITIVE REGULATOR OF CAPSULAR POLYSACCHARIDE SYNTHESIS. RCSA AND RCSB FORM A COMPLEX TO PROMOTE TRANSCRIPTION OF THE GENES FOR CAPSULE SYNTHESIS.  
CC -!- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL REGULATORS.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; X58707; CAA41544.1; -  
PIR; S17701; S17701.  
InterPro; IPR000792; HTH\_LuxR.  
Pfam; PF00196; Gere; 1.  
PRINTS; PR00038; HTHLUXR.  
ProDom; PD000307; HTH\_LuxR; 1.  
SMART; SM00421; HTH\_LuxR; 1.  
PROSITE; PS00622; HTH\_LuxR\_FAMILY; 1.  
KW Transcription regulation; DNA-binding; Activator.  
FT DNA\_BIND 159 178 H-T-H MOTIF (BY SIMILARITY).  
SQ SEQUENCE 211 AA; 24238 MW; 5298CE6C21E6A4A7 CRC64;

Query Match  
Best Local Similarity 6.2%; Score 6; DB 1; Length 211;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TLSQTE 21  
Db 141 TLSQTE 146

RESULT 29

NIA2\_MAIZE  
ID NIA2\_MAIZE STANDARD; PRT; 231 AA.  
AC P39871;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Nitrate reductase [NAD(P)H] (EC 1.6.6.2) (NR) (Fragment).  
GN NAR.  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Panicoideae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. W64A X WI82E; TISSUE=Root;  
RA Long D.M., Oaks A., Rothstein S.J.;  
RT "Regulation of maize root nitrate reductase mRNA levels."  
RL Physiol. Plantarum 85:561-566(1992).  
CC -!- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.  
CC -!- CATALYTIC ACTIVITY: NAD(P)H + nitrate -> NAD(P)(+) + nitrite + H(2)O.  
CC -!- COFACTOR: REQUIRES FAD, A HEME GROUP (CALLED CYTOCHROME B-557) AND ONE MOLYBDENUM ATOM.  
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -!- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE N-TERMINAL DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 CYTOCHROME B5 FAMILY, HEME-BINDING DOMAIN.  
CC -!- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE C-TERMINAL DOMAIN.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; X64446; CAA45776.1; -  
PIR; S24544; S24544.  
HSSP; P17571; 2CND.  
MaizeDB; 25891; -  
InterPro; IPR001199; Cyt\_B5.  
InterPro; IPR001834; Cyt\_B5\_reductase.  
InterPro; IPR000572; Euk\_Mb\_oxred.  
InterPro; IPR001433; Oxred\_FAD/NAD(P).  
InterPro; IPR001230; Prenyl\_site.  
Pfam; PF00175; NAD\_binding; 1.  
Pfam; PF00970; FAD\_binding\_6; 1.  
PROSITE; PS00191; CYTOCHROME\_B5\_1; PARTIAL.  
PROSITE; PS02555; CYTOCHROME\_B5\_2; PARTIAL.  
PROSITE; PS00559; MOLYBDOPTERIN\_EUK; PARTIAL.  
KW Oxidoreductase; Flavoprotein; FAD; NAD; NADP; Heme; Molybdenum;  
KW Nitrate assimilation; Multigene family.  
FT NON\_TER 1  
SQ SEQUENCE 231 AA; 26254 MW; 9864B425C0ED45F4 CRC64;

Query Match  
Best Local Similarity 6.2%; Score 6; DB 1; Length 231;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSET 16  
Db 193 LGDSET 198

RESULT 30  
MTMU\_MYCSP  
ID MTMU\_MYCSP STANDARD; PRT; 233 AA.  
AC P43641;  
DT 01-NOV-1995 (Rel. 32, Created)



DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Modification methylase Muni (EC 2.1.1.72) (Adenine-specific  
 DE methyltransferase Muni) (M.Muni).  
 GN MUNIM.  
 OS Mycoplasma sp.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=2108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94237472; PubMed=8181741;  
 RA Siksnys V., Zareckaja N., Vaisvila R., Timinskas A., Stakenas P.,  
 RA Butkus V., Janulaitis A.;  
 RT "CAATG-specific restriction-modification muni genes from Mycoplasma:  
 RT sequence similarities between R.Muni and R.EcorI.";  
 RL Gene 142.1-8(1994).  
 CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE  
 CC CAATG, CAUSES SPECIFIC METHYLATION ON A-3 ON BOTH STRANDS, AND  
 CC PROTECTS THE DNA FROM CLEAVAGE BY THE MNI ENDONUCLEASE.  
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-  
 CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X76192; CAA53786.1; -  
 DR REBASE; 3449; M.Muni.  
 DR InterPro; IPR002052; N6\_Mtase.  
 DR PROSITE; PS00092; N6\_MTASE; 1.  
 KW Transferase; Methyltransferase; Restriction system.  
 SQ SEQUENCE 233 AA; 26933 MW; 0BE21A05BB2A3A2E CRC64;  
 Query Match 6.2%; Score 6; DB 1; Length 233;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 TLSQTE 21  
 Db 142 TLSQTE 147  
 RESULT 31  
 A29B\_DROSI STANDARD; PRT; 234 AA.  
 ID A29B\_DROSI  
 AC Q9U968;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Accessory gland protein Acp29Ab precursor.  
 GN ACP29AB.  
 OS Drosophila simulans (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7240;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S9;  
 RX MEDLINE=99282496; PubMed=10353898;  
 RA Aguade M.;  
 RT "Positive selection drives the evolution of the Acp29Ab accessory  
 RT gland protein in Drosophila.";  
 RL Genetics 152:543-551(1999).  
 CC -!- FUNCTION: RESPONSIBLE FOR PHYSIOLOGICAL AND BEHAVIORAL CHANGES IN  
 CC MATED FEMALE FLIES.  
 CC -!- SUBCELLULAR LOCATION: Secreted (Probable).  
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AJ240552; CAB53227.1; -  
 DR HSP; P16581; IESL.  
 DR FlyBase; FBgn0027418; Dsim\Acp29Ab.  
 DR InterPro; IPR001304; LECTIN\_C.  
 DR Pfam; PF00059; lectin\_c; 1.  
 DR SMART; SM0034; CLECT; 1.  
 DR PROSITE; PS00615; C\_TYPE\_LLECTIN\_1; FALSE\_NEG;  
 DR PROSITE; PS50041; C\_TYPE\_LLECTIN\_2; 1.  
 KW Behavior; Lectin; Signal.  
 FT SIGNAL 1 21  
 FT CHAIN 22 234  
 FT DOMAIN 137 228  
 FT CARBOHYD 29  
 FT CARBOHYD 61  
 FT CARBOHYD 127  
 FT CARBOHYD 164  
 FT CARBOHYD 234 AA; 26915 MW; EE9C556E02EFED98 CRC64;  
 SQ SEQUENCE 234 AA; 26915 MW; EE9C556E02EFED98 CRC64;  
 Query Match 6.2%; Score 6; DB 1; Length 234;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 72 VTCREM 77  
 Db 137 VTCREM 142  
 RESULT 32  
 BIOD\_YEAST STANDARD; PRT; 237 AA.  
 ID BIOD\_YEAST  
 AC P53630;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Dethiobiotin synthetase (EC 6.3.3.3) (Dethiobiotin synthase) (DTB  
 DE synthetase) (DTBS).  
 GN BIO4 OR YNR057C OR N3506.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FL100;  
 RA Philip V., Jeltsch J.M., Lemoine Y.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Dueterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D., Hilbert H.,  
 RA Moestl D.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- CATALYTIC ACTIVITY: ATP + 7,8-diaminononanoate + CO(2) = ADP +  
 CC phosphate + dethiobiotin.  
 CC -!- COFACTOR: MAGNESIUM.  
 CC -!- PATHWAY: Bioconversion of pimelate into dethiobiotin.  
 CC -!- SIMILARITY: BELONGS TO THE DETHIOBIOTIN SYNTHETASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

```
DR EMBL; U53467; AAB63971.1; -
DR EMBL; Z71672; CAA96339.1; -
DR SGD; S0005340; BIO4.
DR InterPro; IPR004472; BioD.
DR TIGRFAMs; TIGR00347; bioD; 1.
KW Biotin biosynthesis; Ligase; Magnesium; ATP-binding.
FT NP_BIND 18 26 ATP (BY SIMILARITY).
SQ SEQUENCE 237 AA; 26257 MW; F147BF18DA40735 CRC64;

Query Match
Best Local Similarity 6.2%; Score 6; DB 1; Length 237;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 GDSETL 17
Db 53 GDSETL 58

RESULT 33
CW14_YEAST
ID CW14_YEAST STANDARD; PRT; 238 AA.
AC O13547;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Covalently-linked cell wall protein 14 precursor (Inner cell wall
protein).
DE CCW14 OR ICWP OR SSR1 OR YLR391W-A.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=97234625; PubMed=9079899;
RA Moukadiri I., Armero J., Abad A., Sentandreu R., Zueco J.;
RT "Identification of a mannoprotein present in the inner layer of the
cell wall of Saccharomyces cerevisiae.";
RL J. Bacteriol. 179:2154-2162(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albertmann K., Andre B., Ansoerge W.,
RA Benes V., Bruckner M., Delius H., Dubois E., Dusterhoft A.,
RA Entian K.D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Kottler P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Mostl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:87-90(1997).
RN [3]
RP SEQUENCE OF 23-35, AND CHARACTERIZATION.
RX MEDLINE=99255531; PubMed=10322008;
RA Misa V., Ecker M., Strahl-Bolsinger S., Nimtz M., Lehle L., Tanner W.;
RT "Deletion of new covalently linked cell wall glycoproteins alters the
electrophoretic mobility of phosphorylated wall components of
Saccharomyces cerevisiae.";
RL J. Bacteriol. 181:3076-3086(1999).
CC -!- FUNCTION: COMPONENT OF THE INNER LAYER OF THE CELL WALL.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- PTM: Extensively O-glycosylated.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19729; AAB82348.1; -
DR COMPLYEAST-2DPAGE; O13547; -
DR SGD; S0006429; CCW14.
KW Cell wall; Glycoprotein; GPI-anchor; Membrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 220 COVALENTLY-LINKED CELL WALL PROTEIN 14.
FT PROPEP 221 238 REMOVED IN MATURE FORM (POTENTIAL).
FT DOMAIN 77 218 SER-RICH.
FT CARBOHYD 87 87 N-LINKED (GLCNAC...) (POTENTIAL).
FT LIPID 220 220 GPI-ANCHOR (POTENTIAL).
SQ SEQUENCE 238 AA; 23268 MW; 450042DC144C12B8 CRC64;

Query Match
Best Local Similarity 6.2%; Score 6; DB 1; Length 238;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ASLGDS 14
Db 88 ASLGDS 93

RESULT 34
YCND_BACSU
ID YCND_BACSU STANDARD; PRT; 249 AA.
AC .P94424;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical oxidoreductase ycnD (EC 1.-.-.-).
GN YCND.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=97124189; PubMed=8969502;
RA Yamane K., Kumano M., Kurita K.;
RT "The 25 degrees-36 degrees region of the Bacillus subtilis
chromosome: determination of the sequence of a 146 kb segment and
identification of 113 genes.";
RL Microbiology 142:3047-3056(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serrori P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
```

Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A., "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*," Nature 390:249-256(1997).

-!- COFACTOR: FMN (BY SIMILARITY).

-!- SIMILARITY: BELONGS TO THE FLAVIN OXIDOREDUCTASE FRP FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; D50453; BAA09018.1; -. DR EMBL; Z99106; CAB12194.1; -. DR HSSP; Q56691; IBKJ. DR Subtilist; BG12040; ycnD. DR InterPro; IPR000415; Nitroreductase. DR Pfam; PF00881; Nitroreductase; 1. DR Hypothetical protein; Oxidoreductase; Flavoprotein; FMN; KW Complete proteome. KW SEQUENCE 249 AA; 27867 MW; 691BEAE44234FA59 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 249; Best Local Similarity 100.0%; Pred. No. 64; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKK 30 Db 54 KERKKK 59

RESULT 35

AGL6\_ARATH ID AGL6\_ARATH STANDARD; PRT; 252 AA.

AC P29386; DT 01-DEC-1992 (Rel. 24, Created) DT 01-FEB-1996 (Rel. 33, Last sequence update) DT 15-JUN-2002 (Rel. 41, Last annotation update) DE Agamous-like MADS box protein AGL6. GN AGL6 OR AT2G45650 OR F17K2.18. OS Arabidopsis thaliana (Mouse-ear cress). OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; OC eurosids II; Brassicales; Brassicaceae; Arabidopsis. OC NCBI\_TaxID=3702; RN [1] SEQUENCE FROM N.A. RP MEDLINE=91160981; PubMed=1672119; RX Ma H., Yanofsky M.F., Meyerowitz E.M.; RA "AGL1-AGL6, an Arabidopsis gene family with similarity to floral RT homeotic and transcription factor genes."; RL Genes Dev. 5:484-495(1991).

[2] SEQUENCE FROM N.A. RP STRAIN=cv. Columbia; RX MEDLINE=20083487; PubMed=10617197; RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D., Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V., Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L., Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L., Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M., Venter J.C.; RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis RT thaliana."; RL Nature 402:761-768(1999).

-----

-!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.

-!- SUBCELLULAR LOCATION: Nuclear.

-!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION FACTORS.

-!- SIMILARITY: CONTAINS 1 K-BOX DIMERIZATION DOMAIN.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; M55554; AAA79328.1; -. DR EMBL; AC003680; AAC06173.1; -. DR PIR; F39534; F39534. DR HSSP; P11831; ISRS. DR TRANSFAC; T03029; -. DR InterPro; IPR002487; TF\_Kbox. DR InterPro; IPR002100; TF\_MADSbox. DR Pfam; PF00319; SRF-TF; 1. DR Pfam; PF01486; K-box; 1. DR PRINTS; PR00404; MADSDOMAIN. DR SMART; SM00432; MADS; 1. DR PROSITE; PS00350; MADS\_BOX\_1; 1. DR PROSITE; PS00666; MADS\_BOX\_2; 1. DR Transcription regulation; DNA-binding; Nuclear protein. KW DOMAIN 3 57 MADS. FT DOMAIN 95 167 K-BOX. FT SEQUENCE 252 AA; 28744 MW; F763A4A71515CF20 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 252; Best Local Similarity 100.0%; Pred. No. 64; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 LRKKER 27 Db 152 LRKKER 157

RESULT 36

YQEU\_BACSU ID YQEU\_BACSU STANDARD; PRT; 256 AA.

AC P54461; DT 01-OCT-1996 (Rel. 34, Created) DT 01-OCT-1996 (Rel. 34, Last sequence update) DT 15-JUN-2002 (Rel. 41, Last annotation update) DE Hypothetical protein yqeu. GN YQEU. OS Bacillus subtilis. OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus. OC NCBI\_TaxID=1423; RN [1] SEQUENCE FROM N.A. RP STRAIN=168 / JH642; RX MEDLINE=97124195; PubMed=8969508; RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M., RA Kobayashi Y.; RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of RT the Bacillus subtilis genome containing the skin element and many RT sporulation genes."; RL Microbiology 142:3103-3111(1996).

[2] SEQUENCE FROM N.A. RP STRAIN=168 / JH642; RX MEDLINE=97175542; PubMed=9023197; RA Homuth G., Masuda S., Mogk A., Kobayashi Y., Schumann W.; RT "The dnaK operon of *Bacillus subtilis* is heptacistronic."; RL J. Bacteriol. 179:1153-1164(1997).

[3] SEQUENCE FROM N.A. RP STRAIN=168;

RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A., "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*," Nature 390:249-256(1997).

-!- COFACTOR: FMN (BY SIMILARITY).

-!- SIMILARITY: BELONGS TO THE FLAVIN OXIDOREDUCTASE FRP FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; D50453; BAA09018.1; -. DR EMBL; Z99106; CAB12194.1; -. DR HSSP; Q56691; IBKJ. DR Subtilist; BG12040; ycnD. DR InterPro; IPR000415; Nitroreductase. DR Pfam; PF00881; Nitroreductase; 1. DR Hypothetical protein; Oxidoreductase; Flavoprotein; FMN; KW Complete proteome. KW SEQUENCE 249 AA; 27867 MW; 691BEAE44234FA59 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 249; Best Local Similarity 100.0%; Pred. No. 64; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKK 30 Db 54 KERKKK 59

RESULT 35

AGL6\_ARATH ID AGL6\_ARATH STANDARD; PRT; 252 AA.

AC P29386; DT 01-DEC-1992 (Rel. 24, Created) DT 01-FEB-1996 (Rel. 33, Last sequence update) DT 15-JUN-2002 (Rel. 41, Last annotation update) DE Agamous-like MADS box protein AGL6. GN AGL6 OR AT2G45650 OR F17K2.18. OS Arabidopsis thaliana (Mouse-ear cress). OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; OC eurosids II; Brassicales; Brassicaceae; Arabidopsis. OC NCBI\_TaxID=3702; RN [1] SEQUENCE FROM N.A. RP MEDLINE=91160981; PubMed=1672119; RX Ma H., Yanofsky M.F., Meyerowitz E.M.; RA "AGL1-AGL6, an Arabidopsis gene family with similarity to floral RT homeotic and transcription factor genes."; RL Genes Dev. 5:484-495(1991).

[2] SEQUENCE FROM N.A. RP STRAIN=cv. Columbia; RX MEDLINE=20083487; PubMed=10617197; RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D., Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V., Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L., Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L., Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M., Venter J.C.; RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis RT thaliana."; RL Nature 402:761-768(1999).



```
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: BELONGS TO THE UPF0088 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D84432; BAA12467.1; -
DR EMBL; D83717; BAA12079.1; -
DR EMBL; Z99117; CAB14486.1; -
DR Subtilist; BG11645; yqeu.
DR InterPro; IPR004382; Cons_hypoth46.
DR TIGRFAMS; TIGR00046; Cons_hypoth46; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 256 AA; 28802 MW; AAET08FE4283157E CRC64;

Query Match
Best Local Similarity 6.2%; Score 6; DB 1; Length 256;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 KKKRR 33
Db 126 KKKRR 131
|||||

RESULT 37
FPG_VIBCH
ID FPG_VIBCH STANDARD; PRT; 269 AA.
AC Q9KVC5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23) (Fapy-DNA
DE glycosylase).
GN MUTM OR FPG OR VC0221.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
```

```
RN SEQUENCE FROM N.A.
RP STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -!- FUNCTION: This enzyme may play a significant role in processes
CC leading to recovery from mutagenesis and/or cell death by
CC alkylating agents (By similarity).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of DNA containing ring-opened N(7)-
CC methylguanine residues, releasing 2,6-diamino-4-hydroxy-5-(N-
CC methyl)formamidopyrimidine.
CC -!- COFACTOR: Binds 1 zinc ion (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE004112; AAF93397.1; -
DR HSSP; O50606; 1EE8.
DR TIGR; VC0221; -
DR InterPro; IPR000191; Fapy_DNA_glyco.
DR Pfam; PF01149; Fapy_DNA_glyco; 1.
DR ProDom; PD003680; Fapy_DNA_glyco; 1.
DR TIGRFAMS; TIGR00577; fpg; 1.
DR PROSITE; PS01242; FPG; FALSE_NEG.
KW DNA repair; Hydrolase; Glycosidase; Zinc; Zinc-finger;
KW Complete proteome.
FT ZN_FING 244 267
SQ SEQUENCE 269 AA; 30036 MW; 677A73179E5F8887 CRC64;

Query Match
Best Local Similarity 6.2%; Score 6; DB 1; Length 269;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVEVSR 6
Db 6 EVEVSR 11
|||||

RESULT 38
NRTB_PHOLA
ID NRTB_PHOLA STANDARD; PRT; 279 AA.
AC Q51881;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nitrate transport permease protein nrtB.
GN NRTB.
OS Phormidium laminosum.
OC Bacteria; Cyanobacteria; Oscillatoriales; Phormidium.
OX NCBI_TaxID=32059;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=OH-1-P-CL1;
RX MEDLINE=95375238; PubMed=7647306;
RA Merchan F., Kindle K.L., Llama M.J., Serra J.L., Fernandez E.;
RT "Cloning and sequencing of the nitrate transport system from the
RT thermophilic, filamentous cyanobacterium Phormidium laminosum:
RT comparative analysis with the homologous system from Synechococcus
```



```
RT sp. PCC 7942.";
RL Plant Mol. Biol. 28:759-766(1995).
CC -!- FUNCTION: PROBABLY PART OF A HIGH-AFFINITY BINDING-PROTEIN-
CC DEPENDENT TRANSPORT SYSTEM FOR NITRATE; PROBABLY RESPONSIBLE FOR
CC THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -!- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM PERMEASE FAMILY. CYSTW SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; Z19598; CAA79657.1; -
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF00528; BPD_transp; 1.
DR TIGRFAMs; TIGR01183; ntrB; 1.
DR PROSITE; PS00402; BPD_TRANSNIP_INN_MEMBR; FALSE_NEG.
KW Transport; Transmembrane; Nitrate assimilation.
FT TRANSMEM 30 50 POTENTIAL.
FT TRANSMEM 98 118 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 196 216 POTENTIAL.
FT TRANSMEM 217 237 POTENTIAL.
FT TRANSMEM 249 269 POTENTIAL.
SQ SEQUENCE 279 AA; 30204 MW; 5BE86C4B02EFC5F6 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 279;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 EMLTGG 81
Db 224 EMLTGG 229

RESULT 39
NADC_METJA STANDARD; PRT; 283 AA.
AC Q57916;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable nicotinate-nucleotide pyrophosphorylase [carboxylating]
DE (EC 2.4.2.19) (Quinolinolate phosphoribosyltransferase
DE [decarboxylating]) (QAPRTase).
GN NADC OR MJ0493.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcales; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: Nicotinate D-ribonucleotide + diphosphate +
CC CO(2) = pyridine-2,3-dicarboxylate + 5-phospho-alpha-D-ribose 1-
```

```
CC diphosphate.
CC -!- PATHWAY: NAD biosynthesis; aspartate to NAMN; third (last) step.
CC -!- SIMILARITY: BELONGS TO THE NADC/MODD FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U67499; AAB98483.1; -
DR HSSP; O06594; IQPO.
DR TIGR; MJ0493; -
DR InterPro; IPR004393; Nadc.
DR Pfam; PF01729; QRPTase; 1.
DR Pfam; PF02749; QRPTase_N; 1.
DR ProDom; PD003988; QRPTase; 1.
DR TIGRFAMs; TIGR00078; nadc; 1.
KW Pyridine nucleotide biosynthesis; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 283 AA; 31970 MW; BDD118E9CE1401A8 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 283;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 CGIDFI 44
Db 48 CGIDFI 53

RESULT 40
PARB_CHLPN STANDARD; PRT; 286 AA.
AC Q9Z7M0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable chromosome partitioning protein parB.
GN PARB OR CPN0684 OR CP0062.
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RL pneumoniae AR39."
RN Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
```

RT from Japan and CWL029 from USA.";  
RL Nucleic Acids Res. 28:2311-2314(2000).  
CC -!- FUNCTION: INVOLVED IN CHROMOSOME PARTITION. LOCALIZE TO BOTH POLES  
CC OF THE PREDIVISIONAL CELL FOLLOWING COMPLETION OF DNA REPLICATION.  
CC BINDS TO THE DNA ORIGIN OF REPLICATION (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE PARB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

DR EMBL; AE001651; AAD18823.1; -;  
DR EMBL; AE002169; AAF37951.1; -;  
DR EMBL; AP002547; BAA98891.1; -;  
DR PHCI-2DPAGE; Q9Z7M0; -;  
DR TIGR; CP0062; -;  
DR InterPro; IPR004437; ParB\_part.  
DR InterPro; IPR003115; ParBc.  
DR Pfam; PF02195; ParBc; 1.  
DR SMART; SM00470; ParBc; 1.  
DR TIGRFAMS; TIGR00180; parB\_part; 1.  
KW Chromosome partition; DNA-binding; Complete proteome.  
SQ SEQUENCE 286 AA; 32057 MW; 7E0C642FB82DA1B3 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 286;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 59 IQESLL 64  
|||||  
Db 158 IQESLL 163

RESULT 41  
EFTS\_ANASP STANDARD; PRT; 313 AA.  
ID EFTS\_ANASP  
AC Q8YMY3;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Elongation factor Ts (EF-Ts).  
GN TSF OR ALL4791.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
CC -!- FUNCTION: Associates with the EF-Tu.GDP complex and induces the  
CC exchange of GDP to GTP. It remains bound to the aminoacyl-tRNA.  
CC EF-Tu.GTP complex up to the GTP hydrolysis stage on the ribosome.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: BELONGS TO THE EF-TS FAMILY.  
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC EMBL; AP003597; BAB76490.1; -;  
DR InterPro; IPR001816; EF\_TS.  
DR InterPro; IPR000449; UBA\_domain.  
DR Pfam; PF00889; EF\_TS; 1.  
DR Pfam; PF00627; UBA; 1.  
DR TIGRFAMS; TIGR00116; tsf; 1.  
DR PROSITE; PS01126; EF\_TS\_1; 1.  
DR PROSITE; PS01127; EF\_TS\_2; 1.  
KW Elongation factor; Protein biosynthesis; Complete proteome.  
FT SITE 82 85 INVOLVED IN MG++ ION DISLOCATION FROM EF-  
FT TU (BY SIMILARITY).  
SQ SEQUENCE 313 AA; 34275 MW; 234F49543734652A CRC64;

Query Match 6.2%; Score 6; DB 1; Length 313;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EVEVSR 6  
|||||  
Db 280 EVEVSR 285

RESULT 42  
CPPI\_ENTHI STANDARD; PRT; 315 AA.  
ID CPPI\_ENTHI  
AC Q01957;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cysteine proteinase 1 precursor (EC 3.4.22.-).  
GN CPPI OR CPP.  
OS Entamoeba histolytica.  
OC Eukaryota; Entamoebidae; Entamoeba.  
OX NCBI\_TaxID=5759;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=HM-1:IMSS;  
RX MEDLINE=92389981; PubMed=1518524;  
RA Tannich E., Nickel R., Buss H., Horstmann R.D.;  
RT "Mapping and partial sequencing of the genes coding for two different  
RT cysteine proteinases in pathogenic Entamoeba histolytica.";  
RL Mol. Biochem. Parasitol. 54:109-111(1992).  
RN [2]  
RP SEQUENCE OF 4-315 FROM N.A.  
RC STRAIN=HM-1:IMSS;  
RX MEDLINE=91161560; PubMed=1705935;  
RA Tannich E., Scholze H., Nicke R., Horstmann R.D.;  
RT "Homologous cysteine proteinases of pathogenic and nonpathogenic  
RT Entamoeba histolytica. Differences in structure and expression.";  
RL J. Biol. Chem. 266:4798-4803(1991).  
RN [3]  
RP SEQUENCE OF 94-99; 101-110 AND 112-113.  
RC STRAIN=HM-1:IMSS;  
RX MEDLINE=90095985; PubMed=2557443;  
RA Schulte W., Scholze H.;  
RT "Action of the major protease from Entamoeba histolytica on proteins  
RT of the extracellular matrix.";  
RL J. Protozool. 36:538-543(1989).  
CC -!- FUNCTION: INVOLVED IN THE DESTRUCTION OF HUMAN TISSUE BY  
CC E.HISTOLYTICA. CAN ABOLISH ADHESION AND DEGRADE MATRIX PROTEINS  
CC SUCH AS COLLAGEN, LAMININ AND FIBRONECTIN. MAY PLAY A ROLE  
CC IMPORTANT ROLE IN PATHOGENICITY.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

```
DR EMBL; M94162; AAA29090.1; -.
DR EMBL; M64712; AAA29093.1; -.
DR HSSP; P07711; ICJL.
DR MEROPS; C01.050; -.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00139; THIOLESTERASE_CYS; 1.
DR PROSITE; PS00639; THIOLESTERASE_HIS; 1.
DR PROSITE; PS00640; THIOLESTERASE_ASN; 1.
DR Hydrolase; Thiol protease; zymogen; Signal; Multigene family.
KW SIGNAL 1 13 POTENTIAL.
FT PROPEP 14 93 ACTIVATION PEPTIDE.
FT CHAIN 94 315 CYSTEINE PROTEINASE 1.
FT ACT_SITE 118 118 BY SIMILARITY.
FT ACT_SITE 259 259 BY SIMILARITY.
FT ACT_SITE 279 279 BY SIMILARITY.
FT DISULFID 115 161 BY SIMILARITY.
FT DISULFID 152 193 BY SIMILARITY.
FT CONFLICT 4 4 F -> V (IN REF. 2).
FT CONFLICT 269 269 V -> A (IN REF. 2).
SQ SEQUENCE 315 AA; 35056 MW; DF4E3BC795164147 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 315;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 GDSETL 17
Db 137 GDSETL 142

RESULT 43
L767_CAEEL STANDARD; PRT; 316 AA.
ID L767_CAEEL
AC Q09517;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative steroid dehydrogenase Let-767 (EC 1.1.1.-).
GN LET-767 OR C56G2.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Connell M.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP IDENTIFICATION.
RA Kuervers L.M., O'Neill N.J., Baillie D.L.;
RT "Let-767 is a gut-specific dehydrogenase.";
RL (In) Worm Breeder's Gazette 15(3):34(1998).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY. 17-BETA-HSD 3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U23177; AAA64333.2; -.
DR
```

```
DR WormPep; C56G2.6; CE30639.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Steroid biosynthesis; Oxidoreductase; NADP; Multigene family.
FT NP_BIND 47 76 NADP (BY SIMILARITY).
FT ACT_SITE 202 202 BY SIMILARITY.
SQ SEQUENCE 316 AA; 34309 MW; DA3C6377AC4C12CE CRC64;

Query Match 6.2%; Score 6; DB 1; Length 316;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 LRKKER 27
Db 306 LRKKER 311

RESULT 44
H963_HUMAN STANDARD; PRT; 319 AA.
ID H963_HUMAN
AC O14626;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable G protein-coupled receptor H963.
GN H963.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood monocytes;
RX MEDLINE=98036061; PubMed=9370294;
RA Jacobs K.A., Collins-Racie L.A., Colbert M., Duckett M.,
RA Golden-Fleet M., Kelleher K., Kriz R., LaVallie E.R., Merberg D.,
RA Spaulding V., Stover J., Williamson M.J., McCoy J.M.;
RT "A genetic selection for isolating cDNAs encoding secreted proteins.";
RL Gene 198:289-296(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21426338; PubMed=11524702;
RA Joensuu T., Haemaelaeninen R., Yuan B., Johnson C., Tegelberg S.,
RA Gasparini P., Zelante L., Pirvola U., Pakarinen L., Lehesjoki A.-E.,
RA de la Chapelle A., Sankila E.-M.;
RT "Mutations in a novel gene with transmembrane domains underlie Usher
RT syndrome type 3.";
RL Am. J. Hum. Genet. 69:673-684(2001).
CC -1- FUNCTION: ORPHAN RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF002986; AAC51846.1; -.
DR EMBL; AF411849; AAL47763.1; -.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCR_RHODOPSIN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 21 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 22 42 1 (POTENTIAL).
FT DOMAIN 43 48 CYTOPLASMIC (POTENTIAL).
DR
```

```
FT TRANSMEM 49 69 2 (POTENTIAL).
FT DOMAIN 70 89 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 90 110 3 (POTENTIAL).
FT DOMAIN 111 132 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 133 153 4 (POTENTIAL).
FT DOMAIN 154 181 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 182 202 5 (POTENTIAL).
FT DOMAIN 203 224 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 225 245 6 (POTENTIAL).
FT DOMAIN 246 268 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 269 289 7 (POTENTIAL).
FT DOMAIN 290 319 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 3 3 N-LINKED (GLCNAC...)(POTENTIAL).
SQ SEQUENCE 319 AA; 36754 MW; 79B9821C10841114 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 319;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TLSQTE 21
Db 247 TLSQTE 252
```

```
RESULT 45
PRIM_BUCAP
ID PRIM_BUCAP STANDARD; PRT; 319 AA.
AC P32000;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA primase (EC 2.7.7.-) (Fragment).
GN DNAG.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93012960; PubMed=1398077;
RA Lai C.-Y., Baumann P.;
RT "Sequence analysis of a DNA fragment from Buchnera aphidicola (an endosymbiont of aphids) containing genes homologous to dnaG, rpoD, cysE, and secB."
RL Gene 119:113-118(1992).
CC -!- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL RNA PRIMERS FOR THE OKAZAKI FRAGMENTS ON BOTH TEMPLATE STRANDS AT REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS.
CC -!- COFACTOR: BINDS ONE ZINC ION PER MOLECULE (BY SIMILARITY).
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M90644; AAA73233.1; -.
DR PIR; PC1137; PC1137.
DR HSSP; P02923; 1EQN.
DR InterPro; IPR002936; DNAPrim_toprim.
DR Pfam; PF01751; Toprim; 1.
DR SMART; SM00493; TOPRIM; 1.
KW Transferase; DNA replication; DNA-directed RNA polymerase; Primosome;
KW Zinc-finger; Zinc; Metal-binding.
FT NON_TER 1
SQ SEQUENCE 319 AA; 37483 MW; 868521F4AAF58341 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 319;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 25 KERKKK 30
Db 173 KERKKK 178

Search completed: May 11, 2003, 20:12:47
Job time : 22.7434 secs
```



GenCore version 5.1.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 19:33:41 ; Search time 51.5044 Seconds  
(without alignments)  
388.055 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 97  
Sequence: 1 EVEVSRDHASLGDSFTLSQT.....LTGGCLPWATRSHLGRRKCS 97

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0  
Searched: 671580 seqs, 206047115 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

- Database : SPTREMBL\_21.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phage.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_rvirus.\*
  - 16: sp\_bacteriap.\*
  - 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	7.2	36	2 O68941	O68941 rhodospiril
2	7	7.2	42	4 Q96RT9	Q96rt9 homo sapien
3	7	7.2	165	11 Q9CT47	Q9ct47 mus musculu
4	7	7.2	197	10 P93273	P93273 malus domes
5	7	7.2	203	10 Q9FTN4	Q9ftn4 oryza sativ
6	7	7.2	210	5 Q9VR27	Q9vr27 drosophila
7	7	7.2	235	17 Q975Z5	Q975z5 sulfolobus
8	7	7.2	248	10 Q948V1	Q948v1 magnolia pr
9	7	7.2	305	12 Q8QND6	Q8qnd6 ectocarpus
10	7	7.2	310	16 Q8ZPN9	Q8zpn9 salmonella
11	7	7.2	310	16 Q8Z6P0	Q8z6p0 salmonella
12	7	7.2	317	16 Q8R8B4	Q8r8b4 thermoanaer
13	7	7.2	318	10 Q9SL96	Q9sl96 arabisdopsi
14	7	7.2	318	10 Q8VZ43	Q8vz43 arabisdopsi
15	7	7.2	322	4 Q9HDA9	Q9hda9 homo sapien
16	7	7.2	340	17 Q9UYD9	Q9uyd9 pyrococcus

17	7	7.2	374	13	Q9DEV1	Q9dev1 brachydanio
18	7	7.2	375	4	Q9HA26	Q9ha26 homo sapien
19	7	7.2	383	16	Q9PPN7	Q9ppn7 ureaplasma
20	7	7.2	394	10	Q9ZNU3	Q9znu3 arabisdopsi
21	7	7.2	409	16	Q8YJD1	Q8yjd1 brucella me
22	7	7.2	463	16	Q8YW58	Q8yw58 anabaena sp
23	7	7.2	483	5	Q96148	Q96148 plasmodium
24	7	7.2	483	8	Q36097	Q36097 theileria p
25	7	7.2	486	5	Q8WQA7	Q8wqa7 caenorhabdi
26	7	7.2	490	4	Q8WUZ1	Q8wuz1 homo sapien
27	7	7.2	517	4	Q76021	Q76021 homo sapien
28	7	7.2	557	10	Q65538	Q65538 arabisdopsi
29	7	7.2	560	16	Q86320	Q86320 mycobacteri
30	7	7.2	562	5	Q9NEY0	Q9ney0 caenorhabdi
31	7	7.2	628	10	Q81434	Q81434 arabisdopsi
32	7	7.2	648	17	Q28883	Q28883 archaeoglob
33	7	7.2	651	10	Q9LNS7	Q9lns7 arabisdopsi
34	7	7.2	680	5	Q9V6S8	Q9v6s8 drosophila
35	7	7.2	699	5	Q15816	Q15816 dictyosteli
36	7	7.2	703	16	Q9RZF3	Q9rzf3 deinococcus
37	7	7.2	772	4	Q9H1V1	Q9h1v1 homo sapien
38	7	7.2	777	10	Q81868	Q81868 arabisdopsi
39	7	7.2	799	4	Q96N31	Q96n31 homo sapien
40	7	7.2	881	5	Q9BKN8	Q9bkn8 plasmodium
41	7	7.2	1054	10	Q8RYX2	Q8ryx2 oryza sativ
42	7	7.2	1283	5	Q95RH4	Q95rh4 drosophila
43	7	7.2	1286	5	Q21025	Q21025 caenorhabdi
44	7	7.2	1336	5	Q22944	Q22944 caenorhabdi
45	7	7.2	1435	3	Q03291	Q03291 saccharomyc
46	7	7.2	1551	10	Q9SYP0	Q9syp0 arabisdopsi
47	7	7.2	1911	5	Q9W4M7	Q9w4m7 drosophila
48	7	7.2	1987	5	Q9V841	Q9v841 drosophila
49	7	7.2	2434	8	Q9MTH5	Q9mth5 oenothera h
50	7	7.2	3844	5	Q94648	Q94648 plasmodium
51	6	6.2	50	16	Q8VKK7	Q8vkk7 mycobacteri
52	6	6.2	63	17	Q29078	Q29078 archaeoglob
53	6	6.2	74	16	Q8YBH2	Q8ybh2 brucella me
54	6	6.2	80	6	Q9GMM9	Q9gmm9 macaca fasc
55	6	6.2	80	10	Q9SJ65	Q9sj65 arabisdopsi
56	6	6.2	89	16	Q9K5X5	Q9k5x5 bacillus ha
57	6	6.2	90	5	Q9NLX9	Q9nlx9 leishmania
58	6	6.2	103	12	Q89411	Q89411 paramecium
59	6	6.2	104	4	Q9P0E3	Q9p0e3 homo sapien
60	6	6.2	106	16	Q98A87	Q98a87 rhizobium l
61	6	6.2	110	10	Q64396	Q64396 pisum sativ
62	6	6.2	110	17	Q8ZXZ4	Q8zxz4 pyrobaculum
63	6	6.2	111	11	Q9QVN7	Q9qvn7 mus sp. sii
64	6	6.2	116	12	Q8V6V6	Q8v6v6 halovirus h
65	6	6.2	116	16	Q9WYJ5	Q9wyj5 thermotoga

ALIGNMENTS

RESULT 1  
O68941

ID O68941; PRELIMINARY; PRT; 36 AA.  
AC O68941;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Dinitrogenase 3 beta subunit (Fragment).  
GN ANFK.

OS Rhodospirillum rubrum.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;  
OC Rhodospirillum.  
OX NCBI\_TaxID=1085;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Loveless T.M., Bishop P.E.;  
RT "Identification of Genes Unique to Mo-Independent Nitrogenase Systems  
in Diverse Diazotrophs.";  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

```
DR EMBL; AF058778; AAC14327.1; -.
DR InterPro; IPR000510; Oxred_nitrognasel.
DR Pfam; PF00148; oxidored_nitro; 1.
FT NON_TER 36
SQ SEQUENCE 36 AA; 3957 MW; D94F46BCFD437D97 CRC64;

Query Match
Best Local Similarity 7.2%; Score 7; DB 2; Length 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 ELRKKR 27
Db 4 ELRKKR 10

RESULT 2
Q96RT9 PRELIMINARY; PRT; 42 AA.
AC Q96RT9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HBV pres1(21-47) binding protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Yang J.-Y., Wang Y.;
RT "Cloning and identification of a novel Pres1(21-47) binding protein,
RT SlBP.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF271256; AAK58397.1; -.
FT NON_TER 1
SQ SEQUENCE 42 AA; 5559 MW; 52CDC6GAADF9C2FB CRC64;

Query Match
Best Local Similarity 7.2%; Score 7; DB 4; Length 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERK 29
Db 3 RKKERK 9

RESULT 3
Q9CT47 PRELIMINARY; PRT; 165 AA.
AC Q9CT47;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE DNA polymerase epsilon, subunit 2 (Fragment).
GN POLE2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
```

```
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK011194; BAB27458.1; -.
DR MGD; MGI:1197514; Pole2.
FT NON_TER 165
SQ SEQUENCE 165 AA; 19229 MW; 15933826B119C38D CRC64;

Query Match
Best Local Similarity 7.2%; Score 7; DB 11; Length 165;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERK 29
Db 158 RKKERK 164

RESULT 4
P93273 PRELIMINARY; PRT; 197 AA.
AC P93273;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PAFD103 protein (Fragment).
GN PAFD103.
OS Malus domestica (Apple) (Malus sylvestris).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurossids I; Rosales; Rosaceae; Maloideae; Malus.
OX NCBI_TaxID=3750;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GRANNY SMITH;
RA Dong Y.-H., Janssen B.J., Bielecki L.L., Atkinson R.G., Morris B.A.,
RA Gardner R.C.;
RT "Isolating and characterizing genes differentially expressed early in
RT apple fruit development.";
RL J. Am. Soc. Hortic. Sci. 122:752-757(1997).
DR EMBL; U80270; AAC06385.1; -.
FT NON_TER 1
SQ SEQUENCE 197 AA; 21714 MW; 92518666575C3690 CRC64;

Query Match
Best Local Similarity 7.2%; Score 7; DB 10; Length 197;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 30
Db 54 KKERKK 60

RESULT 5
Q9FTN4 PRELIMINARY; PRT; 203 AA.
AC Q9FTN4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE P0005A05.15 protein.
GN P0005A05.15.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
```

```
[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN=CV. NIPPONBARE;
RA  Sasaki T., Matsumoto T., Yamamoto K.;
RT  "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT  clone:P0005A05.";
RL  Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AP002863; BAB16911.1; -.
SQ  SEQUENCE 203 AA; 22741 MW; 1A74C919A6F70852 CRC64;

Query Match          7.2%; Score 7; DB 10; Length 203;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  27 RKKRR 33
Db  186 RKKRR 192

RESULT 6
Q9VR27
ID  Q9VR27 PRELIMINARY; PRT; 210 AA.
AC  Q9VR27;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT  01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE  CG3244 protein.
GN  CG3244.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BERKELEY;
RX  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA  Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA  Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA  Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA  Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA  Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA  Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA  Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA  Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA  Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA  Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA  Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA  Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA  Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA  Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,
RA  Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA  Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA  Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA  Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA  Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA  Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
```

```
RT  "The genome sequence of Drosophila melanogaster.";
RL  Science 287:2185-2195(2000).
DR  EMBL; AE003575; AAF50981.2; -.
DR  FlyBase; FBgn0031629; CG3244.
DR  InterPro; IPR001304; LECTIN_C.
DR  Pfam; PF00059; lectin_c; 1.
DR  SMART; SM00034; CLECT; 1.
DR  PROSITE; PS00615; C_TYPE_LLECTIN_1; UNKNOWN_1.
DR  PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
SQ  SEQUENCE 210 AA; 24368 MW; DA0D011A96D64430 CRC64;

Query Match          7.2%; Score 7; DB 5; Length 210;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  65 CPPSPKE 71
Db  21 CPPSPKE 27

RESULT 7
Q975Z5
ID  Q975Z5 PRELIMINARY; PRT; 235 AA.
AC  Q975Z5;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DT  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  Hypothetical protein ST0283.
GN  ST0283.
OS  Sulfolobus tokodaii.
OC  Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC  Sulfolobus.
OX  NCBI_TaxID=111955;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=JCM 10545 / 7;
RX  PubMed=11572479;
RA  Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA  Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA  Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA  Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA  Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA  Oshima T., Kikuchi H.;
RT  "Complete genome sequence of an aerobic thermoacidophilic
RT  Crenarchaeon, Sulfolobus tokodaii strain7.";
RL  DNA Res. 8:123-140(2001).
DR  EMBL; AP000982; BAB5253.1; -.
DR  InterPro; IPR000836; PRtransferase.
DR  Pfam; PF00156; Pribosyltran; 1.
DR  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 235 AA; 26863 MW; 0F82B46357A350C4 CRC64;

Query Match          7.2%; Score 7; DB 17; Length 235;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  59 IQESLLC 65
Db  35 IQESLLC 41

RESULT 8
Q948V1
ID  Q948V1 PRELIMINARY; PRT; 248 AA.
AC  Q948V1;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DT  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT  01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE  Putative MADS-domain transcription factor MpMADS4 (Fragment).
GN  MpMADS4.
OS  Magnolia praecocissima.
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; Magnoliales; Magnoliaceae; Magnolia.
```

OX NCBI\_TaxID=81865;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ito M., Shiobara S., Tanabe Y., Hasebe M.;  
RT "Organ identities in Magnolian flower."  
RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AB050646; BAB70739.1; -.  
DR InterPro; IPR002487; TF\_Kbox.  
DR InterPro; IPR002100; TF\_MADSbox.  
DR Pfam; PF01486; K-box; 1.  
DR Pfam; PF00319; SRF-TF; 1.  
DR PROSITE; PS50066; MADS\_BOX\_2; 1.  
FT NON\_TER 1  
SQ SEQUENCE 248 AA; 28206 MW; A84E80942F0B6D58 CRC64;

Query Match 7.2%; Score 7; DB 10; Length 248;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 ELRKKR 27  
| | | | |  
Db 136 ELRKKR 142

RESULT 9  
Q8QND6 PRELIMINARY; PRT; 305 AA.  
AC Q8QND6;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE EsV-1-146.  
OS Ectocarpus siliculosus virus.  
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phaeovirus.  
OX NCBI\_TaxID=37665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ESV-1;  
RA Delaroque N., Bothe G., Pohl T., Knippers R., Mueller D.G., Boland W.;  
RT "The complete nucleotide sequence of the Ectocarpus siliculosus virus genome."  
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF204951; AAK14561.1; -.  
SQ SEQUENCE 305 AA; 33947 MW; 1B68698C00CDB999 CRC64;

Query Match 7.2%; Score 7; DB 12; Length 305;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 REMLTGG 81  
| | | | |  
Db 141 REMLTGG 147

RESULT 10  
Q8ZPN9 PRELIMINARY; PRT; 310 AA.  
AC Q8ZPN9;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Putative transcriptional regulator, LysR family.  
GN YDHB OR STM1429.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2."  
RL Nature 413:852-856(2001).  
DR EMBL; AE008762; AAL20351.1; -.  
DR InterPro; IPR000847; HTH\_LysR.  
DR InterPro; IPR005119; LysR\_subst.  
DR Pfam; PF00126; HTH\_1; 1.  
DR Pfam; PF03466; LysR\_substrate; 1.  
DR PRINTS; PR00039; HTHLYSR.  
DR PROSITE; PS00044; HTH\_LYSR\_FAMILY; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 310 AA; 35013 MW; 7C75EBAFB4A2C630 CRC64;

Query Match 7.2%; Score 7; DB 16; Length 310;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSETL 17  
| | | | |  
Db 288 LGDSETL 294

RESULT 11  
Q8Z6P0 PRELIMINARY; PRT; 310 AA.  
AC Q8Z6P0;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Putative transcriptional regulator.  
GN STY1693.  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrell B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18."  
RL Nature 413:848-852(2001).  
DR EMBL; AL627271; CAD01938.1; -.  
DR InterPro; IPR000847; HTH\_LysR.  
DR InterPro; IPR005119; LysR\_subst.  
DR Pfam; PF00126; HTH\_1; 1.  
DR Pfam; PF03466; LysR\_substrate; 1.  
DR PRINTS; PR00039; HTHLYSR.  
DR PROSITE; PS00044; HTH\_LYSR\_FAMILY; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 310 AA; 35058 MW; 41CBCC313DA26E8A CRC64;

Query Match 7.2%; Score 7; DB 16; Length 310;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSETL 17  
| | | | |  
Db 288 LGDSETL 294

RESULT 12



```
Q8R8B4
ID Q8R8B4 PRELIMINARY; PRT; 317 AA.
AC Q8R8B4;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein TTE2096.
GN TTE2096.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013157; AAM25267.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 317 AA; 35043 MW; FB22219729DEF996 CRC64;

Query Match 7.2%; Score 7; DB 16; Length 317;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 IFWILLF 54
Db 16 IFWILLF 22

RESULT 13
Q9SL96
ID Q9SL96 PRELIMINARY; PRT; 318 AA.
AC Q9SL96;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE At2g25670 protein.
GN AT2G25670.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Rensing C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006053; AAD31367.1; -.
SQ SEQUENCE 318 AA; 35128 MW; 2F2E2CD269FFD004 CRC64;

Query Match 7.2%; Score 7; DB 10; Length 318;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 24 KKERKKK 30
Db 160 KKERKKK 166

RESULT 14
Q8VZ43
ID Q8VZ43 PRELIMINARY; PRT; 318 AA.
AC Q8VZ43;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical 35.2 kDa protein.
GN AT2G25670.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene At2g25670 (GI:15225169).";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY065267; AAL38743.1; -.
KW Hypothetical protein.
SQ SEQUENCE 318 AA; 35154 MW; 2F2E2CC278FED004 CRC64;

Query Match 7.2%; Score 7; DB 10; Length 318;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKKK 30
Db 160 KKERKKK 166

RESULT 15
Q9HDA9
ID Q9HDA9 PRELIMINARY; PRT; 322 AA.
AC Q9HDA9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 36.7 kDa protein (Fragment).
GN CATX-11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kairo A., Wang L., Gao Z.Q., Gao Z.P., Boman B.M.;
RT "Isolation of Novel Genes from Human Colonic Epithelial Cells.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083127; AAF98239.1; -.
DR InterPro; IPR002143; Ribosomal_L1.
DR Pfam; PF00687; Ribosomal_L1; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 322 AA; 36676 MW; E2B78F8627E7B52B CRC64;

Query Match 7.2%; Score 7; DB 4; Length 322;
Best Local Similarity 100.0%; Pred. No. 26;
```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 25 KERKKR 31  
|||||||  
Db 225 KERKKR 231

RESULT 16  
Q9UYD9  
ID Q9UYD9 PRELIMINARY; PRT; 340 AA.  
AC Q9UYD9;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE Hypothetical protein PAB1036.  
GN PAB1036.  
OS Pyrococcus abyssi.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=29292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ORSAY;  
RA Heilig R.;  
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ248288; CAB50473.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 340 AA; 40246 MW; 30781FCE6B59C3C5 CRC64;

Query Match 7.2%; Score 7; DB 17; Length 340;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKR 31  
|||||||  
Db 287 KERKKR 293

RESULT 17  
Q9DEV1  
ID Q9DEV1 PRELIMINARY; PRT; 374 AA.  
AC Q9DEV1;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE Lunatic fringe precursor.  
GN LFNG.  
OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Leve C., Gajewski M., Tautz D.;  
RT "Lunatic fringe mRNA."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY007434; AAG12160.1; -.  
DR ZFIN; ZDB-GENE-980605-16; lfrng.  
DR InterPro; IPR003378; Fringe.  
DR Pfam; PF02434; Fringe; 1.  
KW Signal.  
FT SIGNAL 1 35 POTENTIAL.  
FT CHAIN 84 374 LUNATIC FRINGE.  
SQ SEQUENCE 374 AA; 41881 MW; 082F1FD0705B9A8B CRC64;

Query Match 7.2%; Score 7; DB 13; Length 374;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 SLGDSSET 16

Db 51 SLGDSSET 57  
|||||||

RESULT 18  
Q9HA26  
ID Q9HA26 PRELIMINARY; PRT; 375 AA.  
AC Q9HA26;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)  
DE CDNA FLJ12363 fis, clone MAMMA1002380.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=MAMMARY GLAND;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
RA Ninomiya K., Iwayanagi T.;  
RT "NEDO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK022425; BAB14033.1; -.  
DR InterPro; IPR004012; Run.  
DR Pfam; PF02759; RUN; 1.  
SQ SEQUENCE 375 AA; 41861 MW; F7DFBCBB4297CD95 CRC64;

Query Match 7.2%; Score 7; DB 4; Length 375;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KERKKK 30  
|||||||  
Db 254 KERKKK 260

RESULT 19  
Q9PPN7  
ID Q9PPN7 PRELIMINARY; PRT; 383 AA.  
AC Q9PPN7;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE Hypothetical membrane lipoprotein.  
GN UU602.  
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Ureaplasma.  
OX NCBI\_TaxID=134821;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SEROVAR 3;  
RX MEDLINE=20500219; PubMed=11048724;  
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,  
RA Cassell G.H.;  
RT "The complete sequence of the mucosal pathogen Ureaplasma urealyticum."  
RL Nature 407:757-762(2000).  
DR EMBL; AE002158; AAF31016.1; -.  
DR InterPro; IPR001708; 60kDa\_innermeb.  
DR Pfam; PF02096; 60KD\_IMP; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 383 AA; 43714 MW; DD20085C25CDE265 CRC64;

Query Match 7.2%; Score 7; DB 16; Length 383;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 30  
| | | | |  
Db 30 KKERKK 36

RESULT 20  
Q9ZNU3 PRELIMINARY; PRT; 394 AA.  
AC Q9ZNU3;  
DT 01-MAY-1999 (TReMBLrel. 10, Created)  
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Putative extensin.  
GN AT2G46630.

OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=20083487; PubMed=10617197;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
RA Copenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,  
RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
thaliana";  
RN Nature 402:761-768(1999).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=20083487; PubMed=10617197;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
RA Copenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,  
RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
thaliana";  
RN Nature 402:761-768(1999).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=20083487; PubMed=10617197;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
RA Copenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,  
RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
thaliana";  
RN Nature 402:761-768(1999).  
RN [2]

Query Match 7.2%; Score 7; DB 10; Length 394;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 PPSPEV 72  
| | | | |  
Db 124 PPSPEV 130

RESULT 21  
Q8YJDI PRELIMINARY; PRT; 409 AA.  
AC Q8YJDI;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Putative allantoin permease.  
GN BMEI0155.

OS Brucella melitensis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Brucellaceae; Brucella.  
OX NCBI\_TaxID=29459;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;  
RX MEDLINE=20020109; PubMed=11756688;  
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,  
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,  
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,  
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,  
RA Haselkorn R., Kyrpides N., Overbeek R.;  
RT "The genome sequence of the facultative intracellular pathogen  
Brucella melitensis";  
RN Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
DL EMBL; AE009458; AAL51337.1; -.  
DR InterPro; IPR001248; Cyt\_pur\_permease.  
DR Pfam; PF02133; Transp\_cyt\_pur; 1.  
KW Complete proteome.  
SQ SEQUENCE 409 AA; 44996 MW; BDIAC6559DD768B9 CRC64;

Query Match 7.2%; Score 7; DB 16; Length 409;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 SQTELRK 24  
| | | | |  
Db 396 SQTELRK 402

RESULT 22  
Q8YW58 PRELIMINARY; PRT; 463 AA.  
AC Q8YW58;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Hypothetical protein ALL1758.  
GN ALL1758.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
cyanobacterium Anabaena sp. strain PCC 7120.";  
RN DNA Res. 8:205-213(2001).  
DR EMBL; AP003587; BAB73457.1; -.  
DR InterPro; IPR003018; GAF.  
DR InterPro; IPR001932; PP2C-like.  
DR Pfam; PF01590; GAF; 1.  
DR SMART; SM00065; GAF; 1.  
DR SMART; SM00331; PP2C\_SIG; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 463 AA; 52033 MW; E6BCADFDEA8B7AB1 CRC64;

Query Match 7.2%; Score 7; DB 16; Length 463;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 ELRKKER 27  
| | | | |  
Db 199 ELRKKER 205

RESULT 23

O96148  
ID O96148 PRELIMINARY; PRT; 483 AA.  
AC O96148;  
DT 01-MAY-1999 (TReMBLrel. 10, Created)  
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)  
DE Hypothetical 57.8 kDa protein.  
GN PFB0235W.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99021743; PubMed=9804551;  
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,  
RA Koonin E.V., Shallom S., Mason T., Yu K., Fujii C., Pederson J.,  
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perlea M.,  
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,  
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;  
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium  
RT falciparum.";  
RL Science 282:1126-1132(1998).  
DR EMBL; AE001382; AAC71836.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 483 AA; 57785 MW; 9332C1EC9B4DECFB CRC64;  
  
Query Match 7.2%; Score 7; DB 5; Length 483;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 24 KKERKKK 30  
| | | | |  
Db 459 KKERKKK 465  
  
RESULT 24  
Q36097  
ID Q36097 PRELIMINARY; PRT; 483 AA.  
AC Q36097;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).  
GN COI.  
OS Theileria parva.  
OG Mitochondrion.  
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;  
OC Theileria.  
OX NCBI\_TaxID=5875;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=MUGUGA;  
RA MEDLINE=94155854; PubMed=8112303;  
RA Kairo A., Fairlamb A., Gobright E., Nene V.;  
RT "A 7.1 kb linear DNA molecule of Theileria parva has scrambled rDNA  
RT sequences and open reading frames for mitochondrially-encoded  
RT proteins.";  
RL EMBO J. 13:898-905(1994).  
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
CC AND COPPER B (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME  
CC C + 2 H(2)O.  
CC -!- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).  
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
CC INNER MEMBRANE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
DR EMBL; Z23263; CAA80798.1; --

DR InterPro: IPR000883; COX1.  
DR Pfam: PF00115; COX1; 1.  
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
KW Respiratory chain; Transmembrane.  
FT NON\_TER 1  
SQ SEQUENCE 483 AA; 54008 MW; 314438D6EF4CF3D6 CRC64;  
  
Query Match 7.2%; Score 7; DB 8; Length 483;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 40 GIDFIIF 46  
| | | | |  
Db 150 GIDFIIF 156  
  
RESULT 25  
Q8WQA7  
ID Q8WQA7 PRELIMINARY; PRT; 486 AA.  
AC Q8WQA7;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE Y105E8A.17 protein.  
GN Y105E8A.17.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sulston J.E.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases..  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL; AL132876; CAD21666.1; --  
SQ SEQUENCE 486 AA; 55840 MW; A2C9EA88AFA79D0E CRC64;  
  
Query Match 7.2%; Score 7; DB 5; Length 486;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 21 ELRKKER 27  
| | | | |  
Db 235 ELRKKER 241  
  
RESULT 26  
Q8WUZ1  
ID Q8WUZ1 PRELIMINARY; PRT; 490 AA.  
AC Q8WUZ1;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Hypothetical 55.0 kDa protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG;  
RA Strausberg R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC019069; AAH19069.1; --  
DR InterPro: IPR000637; AT\_hook.  
DR InterPro: IPR002143; Ribosomal\_L1.  
DR Pfam; PF00687; Ribosomal\_L1; 1.



DR PRINTS; PR00929; ATHOOK.  
KW Hypothetical protein.  
SQ SEQUENCE 490 AA; 54972 MW; 5E5CDB8A8BC3709 CRC64;

Query Match 7.2%; Score 7; DB 4; Length 490;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKR 31  
| | | | |  
Db 300 KERKKR 306

RESULT 27  
O76021 PRELIMINARY; PRT; 517 AA.  
ID O76021  
AC O76021;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE PBK1 protein.  
GN PBK1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE=99075201; PubMed=9859858;  
RA Huch G., Hohn H.P., Denker H.W.;  
RT "Identification of differentially expressed genes in human trophoblast cells by DDRT-PCR";  
RL Placenta 19:557-567(1998).  
DR EMBL; AJ007398; CAA07491.1; -.  
DR SWISS-2DPAGE; O76021; HUMAN.  
DR InterPro; IPR002143; Ribosomal\_L1.  
DR Pfam; PF00687; Ribosomal\_L1; 2.  
SQ SEQUENCE 517 AA; 58132 MW; BB16E3D96A5EE27A CRC64;

Query Match 7.2%; Score 7; DB 4; Length 517;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKR 31  
| | | | |  
Db 300 KERKKR 306

RESULT 28  
O65538 PRELIMINARY; PRT; 557 AA.  
ID O65538  
AC O65538;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Hypothetical 61.6 kDa protein.  
GN F4D11.190 OR AT4G32610.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Benes V., Rechmann S., Borkova D., Ansoorge W., Hoheisel J.,  
RA Mewes H.W., Mayer K.F.X., Schueller C.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Benes V., Rechmann S., Borkova D., Ansoorge W., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.

RN [3]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AL022537; CAA18600.1; -.  
DR EMBL; AL161581; CAB79978.1; -.  
DR InterPro; IPR003428; MAM33.  
DR Pfam; PF02330; MAM33; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 557 AA; 61561 MW; 90CDA4D869601609 CRC64;

Query Match 7.2%; Score 7; DB 10; Length 557;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KERKKK 30  
| | | | |  
Db 163 KERKKK 169

RESULT 29  
O86320 PRELIMINARY; PRT; 560 AA.  
ID O86320  
AC O86320;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Hypothetical protein RV0976C.  
GN RV0976C OR MTV044.04C OR MT1003.1.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RA Parkhill J.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / OSHKOSH;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains.";  
RT Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
RL EMBL; Z79700; CAB02013.1; -.  
DR EMBL; AE006984; AAK45253.1; -.  
DR TIGR; MT1003; -.  
DR TuberculList; RV0976C; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 560 AA; 59122 MW; 1DBD773C4814AE90 CRC64;

Query Match 7.2%; Score 7; DB 16; Length 560;  
Best Local Similarity 100.0%; Pred. No. 41;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 REMLTGG 81  
|||||  
Db 18 REMLTGG 24

## RESULT 30

Q9NEY0 PRELIMINARY; PRT; 562 AA.

DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Y105E8E.t protein.  
GN Y105E8E.t  
OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.

RA Sulston J.E.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;  
RA none;

RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology."  
RL Science 282:2012-2018(1998).  
DR EMBL; AL132880; CAB60883.2; -.  
DR InterPro; IPR001005; Myb\_DNA\_binding.  
DR SMART; SM00395; SANT; 1.  
SQ SEQUENCE 562 AA; 64912 MW; FDA433D3D888DEF2 CRC64;

## Query Match

Best Local Similarity 7.2%; Score 7; DB 5; Length 562;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 ELRKKER 27  
|||||

Db 235 ELRKKER 241

## RESULT 31

O81434 PRELIMINARY; PRT; 628 AA.

DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE T24H24.5 protein.  
GN T24H24.5 OR AT4G04070.

OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA WASHU;

RT "The A. thaliana Genome Sequencing Project."  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;  
RA Courtney L.; Stoneking T.; Langston Y.; Mead K.;  
RT "The sequence of A. thaliana T24H24."  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;  
RA Waterston R.;  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.

RA Lamar B.; Stoneking T.; Stumpf J.; Mewes H.W.; Lemcke K.;  
RA Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF075598; AAC28203.1; -.  
DR EMBL; AL161499; CAB77875.1; -.

KW Hypothetical protein.

SQ SEQUENCE 628 AA; 68791 MW; B09D244B764722EB CRC64;  
Query Match 7.2%; Score 7; DB 10; Length 628;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TLSQTEL 22  
|||||

Db 222 TLSQTEL 228

## RESULT 32

O28883 PRELIMINARY; PRT; 648 AA.

DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE DNA helicase, putative.  
GN AF1388.

OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;

RA Klenk H.-P.; Clayton R.A.; Tomb J.-F.; White O.; Nelson K.E.;  
RA Ketchum K.A.; Dodson R.J.; Gwinn M.; Hickey E.K.; Peterson J.D.;  
RA Richardson D.L.; Kerlavage A.R.; Graham D.E.; Kyrpides N.C.;  
RA Fleischmann R.D.; Quackenbush J.; Lee N.H.; Sutton G.G.; Gill S.;  
RA Kirkness E.F.; Dougherty B.A.; McKenney K.; Adams M.D.; Loftus B.;  
RA Peterson S.; Reich C.I.; McNeil L.K.; Badger J.H.; Glodek A.; Zhou L.;  
RA Overbeek R.; Gocayne J.D.; Weidman J.F.; McDonald L.; Utterback T.;  
RA Cotton M.D.; Spriggs T.; Artiach P.; Kaine B.P.; Sykes S.M.;  
RA Sadow P.W.; D'Andrea K.P.; Bowman C.; Fujii C.; Garland S.A.;  
RA Mason T.M.; Olsen G.J.; Fraser C.M.; Smith H.O.; Woese C.R.;  
RA Venter J.C.;

RT "The complete genome sequence of the hyperthermophilic, sulphate-  
RT reducing archaeon Archaeoglobus fulgidus."

RL Nature 390:364-370(1997).

DR EMBL; AE01009; AAB89860.1; -.

DR TIGR; AF1388; -.

DR InterPro; IPR004483; put\_DNA\_helic.

DR TIGRfams; TIGR00376; put\_DNA\_helic; 1.

KW Hypothetical protein; Helicase; Complete proteome.

SQ SEQUENCE 648 AA; 73775 MW; B5074507D856A7D5 CRC64;

## Query Match

Best Local Similarity 7.2%; Score 7; DB 17; Length 648;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 GDSETLS 18  
|||||

Db 609 GDSETLS 615